6-2017

# AN EVALUATION AND REVISION OF THE CHILDREN'S BEHAVIOR QUESTIONNAIRE EFFORTFUL CONTROL SCALES

Scott R. Frohn
*University of Nebraska-Lincoln*, scottfrohn@gmail.com

AN EVALUATION AND REVISION OF THE CHILDREN'S BEHAVIOR QUESTIONNAIRE EFFORTFUL CONTROL SCALES

by

Scott R. Frohn

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychological Studies in Education

(Developmental and Learning Sciences)

Under the Supervision of Professors Kathleen Moritz Rudasill and Eric S. Buhs

Lincoln, Nebraska

July, 2017

AN EVALUATION AND REVISION OF THE CHILDREN'S BEHAVIOR QUESTIONNAIRE EFFORTFUL CONTROL SCALES

Scott Raymond Frohn, Ph.D.

University of Nebraska, 2017

Advisors: Kathleen Moritz Rudasill and Eric S. Buhs

The Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) is a popular parent report measure of children's temperament. Effortful control, which refers to processes involved in regulating reactivity to internal and external stimuli, is one factor of temperament measured by the CBQ using five scales tapping multiple dimensions. Numerous studies examining the psychometric properties of the CBQ have shown some problems with the scales, including inconsistent factor structures and measurement noninvariance. Furthermore, the way effortful control is typically defined in the literature, and even according to the CBQ's authors, is inconsistent with how it is actually measured with the instrument. An additional concern which has not yet been addressed is whether the dimensions of effortful control are measured consistently across the ages for which the CBQ is purported to be useful (i.e., are measurement invariant). In Study 1, I evaluate the psychometric properties of the CBQ Effortful Control scales, with particular attention to age-group measurement invariance. Results provide indication that many CBQ items demonstrate various degrees of noninvariance, and provides strong evidence the CBQ Effortful Control scales need revision. In Study 2, with the help of temperament and effortful control content experts, I piloted a large item pool with a sample of parents using an online questionnaire. Using a subsample of pilot participants (*N* = 400) I refined the scales based on a series of decision rules based in Classical Test Theory (CTT) and Confirmatory Factor Analysis (CFA). The scales were further refined on theoretical grounds with the help of content experts. The psychometric properties of the revised and original scales were then compared in CTT and CFA frameworks with the remaining pilot sample (*N* = 272). The revised scales show considerable improvement over the original scales in content, length, and factor structure, and are comparable to the original versions in terms of CTT reliability. Implications, limitations, and future directions are discussed.

# DEDICATION

This dissertation is dedicated to my mother, Sally Frohn.

I love you, and miss you dearly.

AUTHOR'S ACKNOWLEDGEMENTS

I would like to thank Sam Putnam, Masha Gartstein, Carmen González Salinas, Kathleen Moritz Rudasill, and Mary Rothbart for dedicating their time and energy to this project. The revision process would have been impossible without their collaboration, valuable wisdom, hard work, and willingness to let some guy from Nebraska mess with their favorite measure of children's temperament.

I would like to thank my advisors Kathleen Moritz Rudasill and Eric Buhs, as well as Victoria Molfese, for their mentorship over the past five years has helped me become a better writer, thinker, and scholar. I would like to thank Anthony Albano for his suggestions on conducting the pilot. I would also like to thank my committee members for their valuable feedback for final edits and revisions.

I would like to thank my officemate and friend Amanda Prokasky for challenging my ideas, providing valuable feedback, helping me recruit for the pilot study, maintaining the cache of candy in our office, and generally pushing me to be a better and more thoughtful researcher. I would also like to thank Amy Encinger for helping me recruit for my pilot study and wordsmithing my recruitment language.

For allowing my use of their CBQ data in Study 1, I would like to Dr. Robert Cole and Elizabeth Anson at the University of Rochester, Dr. Stephanie Carlson and Dr. Megan Gunnar at the University of Minnesota, Dr. Grayzna Kochanska at the University of Iowa, Dr. Lucy LeMare at Simon Fraser University, Dr. John Bates at the University of Indiana, Dr. Victoria Molfese and Dr. Kathleen Moritz Rudasill at the University of Nebraska-Lincoln, Dr. Leslie Leve and Dr. Philip Fisher at the University of Oregon, and Dr. Sam Putnam at Bowdoin College.

I would like to thank Rebecca Eanes at Positive Parenting, Tracy Cassels at Evolutionary Parenting, Katie at Mama the Reader, Laura Mazza - Mum on the Run, Victoria Molfese, Aaron Walker, Morley Elementary PTO, Cavett Elementary PTO, Calvert Elementary PTSO, Eastridge Elementary PTO, Jennifer Miller and the Ruth Staples Child Development Center, the Early Development and Learning Lab, the UNL department of Educational Psychology, the UNL Department of Child, Youth, and Family Studies, and my friends and countless friends of friends on Facebook for graciously spreading the word about my study. I would also like to thank my participants.

Finally, I would like to thank my wife, Lindsey, for putting up with my absence in the evenings and on weekends so I could work and write. I love you, and I look forward to our post-graduate school adventures.

TABLE OF CONTENTS

LIST OF MULTIMEDIA OBJECTS

# CHAPTER 1

# LITERATURE REVIEW

## Introduction

The Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fischer, 2001) is a widely used parent report measure of child temperament for developmental and educational research. Five CBQ scales are used for measuring effortful control, which refers to children's ability to regulate and control their behavior, emotions, and attention. While the CBQ effortful control scales are intended to be used for children ages 3 to 7 years, it is unclear whether the underlying constructs of effortful control are measured validly and consistently for children across this age range. Furthermore, a number of studies have identified issues with the effortful control scales and how temperament is measured with the CBQ and its shortened variants (Allan et al., 2013; Clark et al., 2016; de la Osa et al., 2014; Kotelnikova et al., 2016). These findings indicate that the CBQ effortful control scales, among others, may need revision. In this chapter, I review literature regarding the structure and measurement of effortful control, the history and construction of the CBQ, and the role of measurement in psychological sciences. I then evaluate the CBQ effortful control scales using a large sample of children ages 3 to 7 years (Chapter 2). In Chapter 3 I detail the revision, pilot, and evaluation of the CBQ effortful control scales. In Chapter 4 I discuss the reliability and validity evidence of the revised scales, examine some broad implications of revising these scales, and posit future directions.

**Temperament and Effortful Control**

Most contemporary temperament research is based on the work of Mary Rothbart who conceptualized temperament as “constitutionally based individual differences in reactivity and self-regulation in response to changes in internal and external stimuli, in the domains of affect, activity, and attention” (Rothbart and Bates, 2006, p. 99). The term “constitutional” is used to highlight the biological foundation of temperament, indicating that temperament is not learned but innate, detectable early in life, and permeates throughout an individual’s interaction with the environment. Reactivity refers to emotional, attentional, motor, and physiological responsiveness to internal and external stimuli (Rothbart & Bates, 2006). Self-regulation, then, reflects the ability to modulate such reactivity by purposefully controlling behavior, attentional focus, and emotions (Rothbart & Bates, 2006).

One emergent theme in literature regarding the nature of temperament is that temperament must be *elicited* (Rothbart & Bates, 2006). This insight is described in the definition of temperament that these traits arise *in response* to changes in internal states or the external environment. For example, a fearful child may not display fear during a pleasant event, as a child that is easily frustrated may not show signs of frustration during a task that is simple to complete, and a child that takes joy in high-intensity activities may or may not derive joy from simply sitting and reading a book. However, in situations that typically elicit fear (e.g., seeing a large spider) or frustration (e.g., being physically restrained), or during intense activities (e.g., riding a bicycle downhill), these children may respond to these events more quickly, more intensely, and for a longer duration than other children who are less fearful, less easily frustrated, or less inclined toward intense activities.

Among the theoretical models that have been used to account for temperamental contributions to concurrent and future adjustment, child-by-environment models of development have much support, both in terms of theory and predictive value. While main-effects and mediation models are useful for detecting some direct and indirect effects, respectively, they fail to capture the interactive contributions of individual and environmental factors on child development, a central focus of temperament research since the 1970s. Thomas and Chess (1977) explicitly acknowledged the importance of both temperament and context in their “goodness of fit” hypothesis, where they highlighted the interactive relations between individual differences and the environment. Thomas and Chess hypothesized that positive development occurs when a child’s temperament “fits” with the environment. For instance, a highly active child may excel in a classroom in which her movement is not restricted, and where the teacher provides her with duties that allow her to move about the room (e.g., handing out materials, picking up assignments). Furthermore, children also play a role in shaping their environment. For example, adaptable children tend to elicit more warm and responsive caregiving (Putnam, Sanson, & Rothbart, 2002), and children high in negative emotionality from lower socioeconomic status (SES) families tend to elicit less supportive parenting, whereas similar children from higher SES backgrounds elicit greater levels of parental support (Paulussen-Hoogeboom, Stams, Hermanns, & Peetsma, 2007). It is generally accepted that temperament interacts with the environment (Rothbart, Ahadi, Hershey, & Fisher, 2001; Rothbart & Bates, 2006; Rueda, 2012), and many contemporary researchers have adopted a child-by-environment approach to studying temperament, finding that the impact of temperament on social and academic outcomes is moderated by environmental or individual factors (Belsky & Pluess, 2009; Hamre & Pianta, 2001; Lengua et al., 2008; Pluess & Belsky, 2010; Rudasill, Gallagher, & White, 2010; Stright, Gallagher, & Kelley, 2008; Vermeulen et al., 2016).

Rothbart has identified three primary dimensions of temperament: negative affectivity, extraversion/surgency, and effortful control (Rothbart, 2007; Rothbart & Bates, 2006; Rothbart et al., 2001). Negative affectivity refers to aspects of negative emotionality, including reactive distress in response to overwhelming sensory stimuli (discomfort), limitations (anger/frustration), and disappointment (sadness), as well as anticipation of distress (fear) and degree to which a child recovers from distress (soothability) (Rothbart & Bates, 2006). Extraversion—also referred to as "surgency"—refers to aspects of positive emotionality and stimuli-seeking behavior, including displays of joy (smiling and laughter), pleasure derived from intense activity (high intensity pleasure), degree of outgoingness (reverse of shyness), excitatory anticipation for desired activities (approach), rate and duration of gross motor movement (activity), and quickness to initiate a response (impulsivity).

The third broad dimension within Rothbart's conceptualization of temperament is that of effortful control, representing processes involved in orienting and regulating reactivity. Effortful control is often described in terms of the components of which it is thought to be comprised. These components include the abilities to (a) suppress or inhibit a dominant response, (b) perform a subdominant response, (c) plan actions, (d) detect errors, (e) focus attention, and (f) shift attention (Rothbart et al., 2001; Rothbart & Bates, 2006; Rueda, 2012). However, effortful control is also defined within the context of the executive attention system, which is thought to underlie effortful control functioning (Posner & Rothbart, 2000; Rothbart, Sheese, & Posner, 2007). Indeed, Rothbart and Bates (2006) regarded effortful control as the "efficiency of executive attention," (p. 126) indicating that the processes involved in effortful control are, in essence, a function of how information is selected for attention and how attention is deployed and sustained (Bush, Luu, & Posner, 2000; Rothbart & Bates, 2006). For instance, consider "Go-No Go" paradigms used in investigations of inhibitory control, such as Simon Says (Kochanska, Coy, & Murray, 2001). In these types of tasks, the study child must inhibit a dominant response (e.g., inhibiting an action when a directive is not prefaced with "Simon Says"), which requires particular attention to cues preceding a directive and sustaining that attentional awareness to cues throughout the testing trials.

Effortful control is often studied as the temperamental contribution to self-regulation (Bell & Deater-Deckard, 2007; Bridgett et al., 2013). While aspects of negative affectivity and extraversion/surgency have some links to internalizing and externalizing behaviors (Eisenberg et al., 2001; Rothbart, Ahadi, & Hershey, 1994; Wolfe et al., 1987), effortful control and self-regulation are frequently and strongly implicated in social and academic success from early childhood through adulthood (Bandura, 2005; Blair, 2002; Calkins, 2007; Eisenberg, Valiente, & Eggum, 2010; Rothbart & Jones, 1998; Willoughby et al., 2011).

While negative affectivity and extraversion/surgency are important to understanding human functioning, the saliency of effortful control to social and academic outcomes of interest to parents, educators, researchers, and policy makers, coupled with psychometric issues relating to factor structure and measurement consistency, it is critical for advancing the scientific understanding of human development to have a valid and reliable measure of effortful control. It is for these reasons I have chosen to focus solely on evaluating and revising current measures of effortful control and not other aspects of temperament. The process presented here, however, will provide a framework through which measures of negative affectivity and extraversion/surgency can be evaluated and revised in the future.

## Effortful Control and Development

Effortful control is implicated in predicting academic and social adjustment, typically with greater effortful control related to more positive developmental outcomes. Researchers typically use the term *effortful control* referring to inhibitory control, attentional focusing, or both aspects of regulation, although conceptually there are other components (Rothbart & Rueda, 2005; Rothbart & Bates, 2006). Kim, Nordling, Yoon, Boldt, and Kochanska (2013) found that attentional focusing predicted future academic performance, and ability to delay and regulate emotions predicted fewer future behavior problems. Effortful control is also related to better academic achievement, and this association may be moderated by greater school engagement. (Iyer, Kochenderfer-Ladd, Eisenberg, & Thompson, 2010). Lengua and colleagues (2008) found that contextual risk factors (e.g., low SES, maternal depression, household and neighborhood quality) were related to the development of internalizing and externalizing problems only for grade school children low in effortful control; children high in effortful control were buffered from deleterious effects of developmental risk. There is also evidence that children high in effortful control demonstrate better social functioning, have more positive relationships with peers and teachers, and are less likely to be victimized by peers (Eisenberg et al., 2003; Iyer et al., 2010; Valiente, Lemery-Chalfant, Swanson, Reiser, 2008; Zhou et al., 2004). Effortful control also allows for attentional flexibility, which is associated with the internalization of moral principles and feelings of guilt and shame, because flexibly controlling and shifting attention allows children to notice the feelings of themselves and others, and helps link a child's feelings to their own actions and moral principles (Derryberry & Reed, 1994, 1996).

Compared to other aspects of temperament, effortful control develops dramatically over the early years of life (Posner & Rothbart, 2007) and demonstrates substantial heterotypic continuity over time. The term heterotypic continuity here means that while the underlying concept of effortful control remains stable, how it is expressed and how we measure it may change over time. Thus, effortful control doesn't manifest the same way across the lifespan. For instance in infancy, duration of orienting and soothability are thought of as nascent aspects of effortful control and self-regulation, and are measured using behavioral tasks or questionnaires, such as the Infant Behavior Questionnaire (IBQ; Garstein & Rothbart, 2003).

Around 12 months of age, infants are able to inhibit a dominant response (reaching for a toy in the same location it had consistently appeared before) in favor of a subdominant response (reaching for a toy placed in a new location) (Diamond, 1991). At 24 months, children seem to have some difficulty acting on conflicting information. Gerardi-Caulton (2000) presented children with a spatial conflict task at age 24 months and again at 36 months. In this task, children are set in front of a screen with two response buttons, one on their left and one on their right. Each button has a different picture on it, and when a picture appears on the screen, the child must press the button with the matching picture. The conflicting situation arises when a picture appears on the side of the screen opposite the corresponding button, thus the child must both inhibit a dominant response (to press the button on the same side) and activate a subdominant response (to press the button on the opposite side). At age 24 months, children are generally unable to perform this task, but by 36 months, they nearly all do so with high accuracy, albeit with a slight delay. Children also seem to develop the capacity to detect and correct errors in task performance (measured by response times following an error) between 24 months and 42 months of age (Jones, Rothbart, & Posner, 2003; Rothbart et al., 2003). Attentional focusing also seems to increase dramatically between ages 4 and 7 years. Performance on the flanker task, which measures children's ability to focus on a target while simultaneously ignoring irrelevant stimuli, improves through age 7, where after performance remains relatively stable though adulthood (Rueda et al., 2004).

Putnam, Rothbart, and Gartstein (2008) examined the stability of temperament traits from infancy through about age five using the IBQ, the Early Childhood Behavior Questionnaire (ECBQ), and the CBQ. The authors found considerable temporal relations between all aspects of infant temperament and low intensity pleasure in early childhood (18-36 months), and between activity, frustration, fear, falling reactivity and several aspects of effortful control in the preschool years (36-59 months). Considering the items on each of these scales, it is clear that how effortful control is currently measured varies dependent on the child's age. For instance, items measuring the maintenance of attentional focus in infancy (e.g., "How often during the last week did the baby repeat the same movement with an object for two minutes or longer?") are contextually different from items measuring the trait in early childhood (e.g., “While looking at picture books on his/her own, how often did your child stay interested in the book for 5 minutes or less?”). Although the underlying construct of attentional focusing—the content of the items—remains the same, how this trait is measured is age and context dependent.

A common theme across studies of the development of effortful control is that performance on effortful control tasks improves over time, but to establish developmental trends of effortful control more difficult tasks requiring multiple components of regulation are needed. For instance, Carlson (2005) found that from ages 4 to 6, two effortful control tasks that children are least likely to pass are Simon Says and Disappointing Gift. Performance on these two tasks increased over time, however, even by age 6 children were less than 50% likely to pass them. In Simon Says, the experimenter instructs the child to perform a commanded action only when preceded by “Simon Says.” Unlike other Go / No Go tasks, when giving the command the experimenter performs the actions regardless of whether or not “Simon Says” was stated. Thus, the task requires both inhibitory control (not performing an action) and attentional focusing (ignoring the irrelevant stimuli of the experimenter’s actions). In Disappointing Gift, the child opens a much anticipated gift to discover only a dull wooden chip. The task is coded for the child’s ability to withhold a negative response and display a positive response, which requires inhibitory control, context monitoring—sometimes conceptualized as a component of effortful control (see: Rothbart & Rueda, 2005)—and perceptual sensitivity (noticing and interpreting the social cues) which are key components of social information processing and in turn social competence (Crick & Dodge, 1994; Rhoades, Greenberg, & Domitrivich, 2009).

Although Carlson (2005) expressly identified the aforementioned tasks as indicators of executive functioning, they clearly demand regulation and are sometimes used for assessing effortful control (see: Kochanska, Coy, & Murray, 2001). Additionally, a major component of executive functioning is working memory (Zhou, Chen, & Main, 2012). One working memory task, backward digit span, administered to children ages 3 to 6 years showed a dramatic increase in probability of passing from age 3 (<10% likely), to age 4 (< 30% likely), and to age 5 and 6 (<90% likely) (Carlson, 2005). This evidence indicates that even with large gains in working memory, children still struggle with tasks involving multiple aspects of effortful control, and that these two related constructs—working memory and effortful control—can be measured separately and may have different developmental trajectories.

## The Children's Behavior Questionnaire

Used since the early 1990's, the most common measure of temperament for children age three to seven is the CBQ. Although the first reference to the CBQ in temperament literature emerged in 1991 (Goldsmith & Rothbart, 1991), the first publication of an in-depth investigation into its psychometric properties and validity evidence was not available until the early 21st century (Rothbart, Ahadi, Hershey, & Fisher, 2001). In this publication, Rothbart and colleagues indicated that their original intent for the instrument was to develop items that could apply to children up to age eight, however the "initial scale development was limited to the 3-7 year age range" (p. 1395) and it is mainly this population to which the CBQ has been applied ever since.

Rothbart and colleagues (2001) adopted Fiske's (1971) rational approach to scale construction by establishing theoretically-derived components of temperament (e.g., effortful control) and breaking down these components into distinct subcomponents (e.g., inhibitory control, attentional focusing, etc.). Items were drawn from two sources: existing questionnaires and parent interviews. First, items were adapted from existing measures of adult temperament (Psychological Reactions Questionnaire; Derryberry & Rothbart, 1988) and infant temperament (Infant Behavior Questionnaire; Rothbart, 1981) for application with young children. Next, a sample of parents of young children were interviewed, providing information comparing individual differences between siblings and which types of situations elicit certain responses. Items constructed from interviews were combined with adapted items, and a small sample of parents was tasked with reviewing the questionnaire and providing suggestions for edits and additional items. CBQ item candidates were then administered to parents of over 250 children ages 3 to 7 years. Item and scale statistics were then computed, and items were eliminated if they exhibited low item-total correlations (<.20) with the intended subscale or relatively higher item-total correlations with another subscale.

The resulting 195 items became what is now known as the standard form of the CBQ, which consists of 16 subscales comprised of 12 to 13 items each. In addition to the four common EC subscales, a fifth subscale of Attentional Shifting (5 items) still appears on the standard form, yet is not represented in shorter versions of the CBQ, nor is typically incorporated into factor scores of EC. While Attentional Focusing and Attention Shifting are related constructs—attention shifting refers to flexibly switching attention to different tasks—and highly correlated (e.g., Derryberry & Rothbart, 1988), the items did not perform well as a single scale. Rothbart and colleagues (2001) acknowledged shortcomings of the Attention Shifting subscale, and indicated that a full scale was being constructed in their laboratory (although it is unclear if this scale has indeed been completed).

Because the length of the 195-item standard form may be prohibitive in some instances, such as a large-scale study with many other parent report measures, the CBQ Short Form (CBQ-SF; Putnam & Rothbart, 2006) was created from CBQ items to approximate the same scales at a considerably shorter length (94 items). Items were selected for inclusion in the CBQ-SF based on two primary criteria: high item-total correlations while retaining the breadth of content of the original CBQ. A third form, the CBQ Very Short Form (CBQ-VSF; Putnam & Rothbart, 2006) was created using 36 items from the CBQ-SF to produce orthogonal factor scores for the three dimensions of the CBQ: negative emotionality, extraversion/surgency, and EC. Items were selected based on high item-total correlations with the parent factor and low correlations with the other factors. The VSF EC factor is comprised of three items each from attentional focusing, inhibitory control, low intensity pleasure, and perceptual sensitivity scales. However, the practice of using item-total correlations as a primary selection criterion has been criticized, as the derived factors are strongly dependent on the item properties (which in turn are sample dependent), and therefore the factor structures may be inaccurate and less likely to be replicated (Kotelnikova et al., 2016). Indeed, some exploratory factor analyses of the CBQ, CBQ-SF, and CBQ-VSF have failed to replicate item-level and scale-level factor structures (Allan et al., 2013; de la Osa et al., 2014; Kotelnikova et al., 2016), suggesting that the items and scales may need to be revised.

The effortful control scales contain a broad representation of content related to the corresponding dimension. The attentional focusing and attention shifting scales show some overlap, with attentional focusing items relating to sustaining attention, task completion, attention shifting, and distractedness, while the attention shifting scale contains items relating to attention flexibility, ability to stop an activity when asked, and concentration. Inhibitory control incorporates content such as controlling motor movement, following directions, planning, lowering voice, stopping a prohibited act, and being careful. Low intensity pleasure items involve mainly enjoyment of various pleasant activities and events, such as enjoyment from words, song, or sounds, enjoying touch, and enjoying books. Perceptual sensitivity content relates to noticing appearances, sounds, textures, smells and tastes.

When comparing the content of the effortful control scales to how effortful control is defined in the literature, however, there are a few gaps in content coverage. As previously mentioned, effortful control is described as the ability to (a) inhibit a dominant response, (b) perform a subdominant response (activational control), (c) plan actions, (d) detect errors, (e) focus attention, and (f) shift attention (Rothbart et al.,2001; Rothbart & Bates, 2006). While CBQ inhibitory control and attentional focusing scales explicitly capture definitional aspects of effortful control, the attention shifting scale is very short, and is not even included in the CBQ-SF, yet it is expressly defined as a component of effortful control. Perceptual sensitivity may be construed as capturing aspects of error detection (e.g., CBQ 98 *"is quickly aware of some new item in the living room"*), and six items from this scale may be considered to cover that content. Planning actions is, in contrast, poorly represented in current CBQ content, with only one related item appearing on the inhibitory control scale (CBQ 63 *"Prepares for trips and outings by planning things s/he will need"*). Performing a subdominant response has some representation in the inhibitory control scale, and it could be argued that inhibitory control and activation control are directly related to one another, as withholding an instinctive reaction (e.g., not running into the street after a stray ball) usually involves also doing something you may not automatically want to do (e.g., waiting at the curb and look both ways before retrieving the ball).

Thus, there exists inconsistency in how effortful control is defined and how it is measured with the CBQ. Furthermore, given the heterotypic continuity and development of effortful control it is possible, if not likely, that the CBQ effortful control items and scales demonstrate differential functioning between ages three and seven. For instance, certain items may be more applicable to younger children than to older children, such as the low intensity pleasure item “enjoys taking warm baths,” as older children are more likely to take showers. If our measurement of effortful control is not consistent over time (i.e., demonstrates longitudinal measurement noninvariance), then we cannot adequately compare scores of younger children with older children because how the construct is represented would not hold the same meaning for these groups, nor can we accurately investigate developmental changes in effortful control because the scale of effortful control would be partly influenced by incongruent measurement by age.

To determine if differential functioning exists, a systematic evaluation of the CBQ effortful control scales needs to be conducted, comparing scale performance for younger children and older children. No study to date has evaluated the psychometric properties of the CBQ across age, although there have been several examinations of the psychometric properties of the CBQ, CBQ-SF, and CBQ-VSF (Clark et al., 2016; de la Osa et al., 2014;

Frohn, Prokasky, & Putnam, 2016; Kotelnikova et al., 2016), as well as a multimethod examination of gender and ethnicity differences effortful control (Sulik et al., 2010).

de la Osa and colleagues (2014) examined the factors structure of the CBQ-SF and VSF in a Spanish sample of three-year-olds. They conducted an EFA using CBQ-SF scale scores, and found them to load well on the predicted factors. However, a CFA on the VSF items using previously published factor structures (Putnam & Rothbart, 2006) failed to produce adequate fit statistics. A principal axis factor analysis with oblique rotation produced a VSF factor structure similar to other studies, yet some items had poor factor loadings (< .30) on the EC factor. Using a sample of four-year-olds, Sulik et al (2010) tested measurement invariance by child gender and ethnicity for a composite effortful control factor, which included a combination of behavioral tasks and CBQ EC (attentional focusing and inhibitory control subscales). The authors found configural, metric, and scalar invariance for the CBQ EC composite held across ethnicities (White, African American, and Hispanic), yet scalar noninvariance was detected across gender, with girls having a higher CBQ EC intercept than boys. Gender noninvariance for the CBQ EC scales at the factor level was also found by Clark and colleagues (2016), but only for the low intensity pleasure scale. In another study, Kotelnikova and colleagues (2016) conducted an exploratory factor analysis (EFA) on the full CBQ scale and identified 15 lower-order factors, only one of which resembled an original CBQ EC scale (attentional focusing). Finally, Frohn, Prokasky, and Putnam (2016) found many items on the CBQ EC scales to have response options with infrequent use, particularly at the low-end of the scale, and recommended that the CBQ rating scale should be reduced.

These findings together suggest uncertainty regarding the efficacy of the CBQ EC scales and items, and indicate a systematic psychometric evaluation and revision may be necessary and desirable. Yet before addressing the CBQ, we first need to understand the nature of psychological measurement, how it contributes to psychological science, and the processes through which measures are constructed, evaluated, and revised. Once we have established guidelines, we can then pursue actions to improve the CBQ EC scales.

## Measurement and Scale Evaluation

One of the earliest definitions of measurement comes from the psychologist S. S. Stevens, who stated "that measurement, in the broadest sense is defined as the assignment of numerals to objects or events according to rules" (1946, p. 677). This definition has since been expanded, with the recognition that measurement involves assigning numerals according to rules *based on properties of objects or events* (Duncan, 1984). Within psychological and educational testing the *object* of measurement is often the individual, and the qualities which we seek to measure (e.g., intelligence, temperament, verbal ability) are typically unobservable. This then presents a problem to psychologists, educators, and researchers: how do we measure that which we cannot observe?

The role of measurement and measurement theory in the development of educational and psychological tests arises from this need to quantify that which cannot be directly observed. Unobservable attributes theoretically exist (e.g., the ability to regulate behavior) and are referred to as *latent variables* or *constructs*; their existence can only be inferred by the indirect measurements we are able to observe. For instance, as mentioned above, effortful control typically refers to an individual's degree of perceptual sensitivity and ability to regulate one's behavior and attention (Rothbart & Bates, 2006). This is our description of the *nature* of the latent variable (de Ayala, 2009), and informs our approach to collecting observations and measuring effortful control as a construct. We may choose to measure effortful control, for example, via parent report questionnaires (e.g., IBQ, ECBQ, CBQ) or lab tasks for measuring self-regulation (e.g., Simon Says). The approach we select to measure effortful control becomes our *operational definition* of the construct, and the observational data we collect is then *scaled* corresponding to values of the latent variable to facilitate interpretation (de Ayala, 2009).

How we develop, evaluate, and use an instrument to measure a given construct is governed by measurement theory. Hubley and Zumbo (2013) identify at least six interrelated measurement theories useful in test/scale development and application, including classical test theory (CTT), factor analysis (e.g., CFA), and item-response theory (IRT). Psychological and educational measures are typically designed to quantify an unobservable attribute of an individual by assigning them a value or score on a given scale corresponding to the level of the attribute the individual possesses. Given that a measurement instrument is used as intended (i.e., with a particular population and for a particular purpose), a good instrument produces scores which discriminate well between individuals with similar yet different levels of the attribute to be measured. The characteristics of the items and the instrument itself provide insight into how accurately and consistently scores are assigned, and how these characteristics are defined and evaluated are governed by measurement theory. Most measurement frameworks allow for indices of item discrimination (how well an item differentiates individuals at different levels of the measured attribute) and difficulty. However, theoretical frameworks differ in how these indices are conceptualized and parameterized, and therefore have different criteria for retaining or omitting items. For instance, within CTT, item selection may be based on parameters that index how well items relate to themselves (internal consistency), how an item relates to the total score (discrimination), and mean item score (difficulty). Item selection within IRT, on the other hand, may be based on item *information* and the range of ability levels across which an item discriminates (Edelen & Reeve, 2013; Penfield, 2013).

It should be mentioned that there exists a distinction between measurement theory and measurement models. Measurement theory refers to framework used to link observations to unobserved constructs, and therefore facilitate test construction and evaluation (Hambleton & Jones, 1993). Models, on the other hand, are the application of measurement theory used to test and explain the relations among variables. The accuracy and appropriateness of a model is evaluated by its fit to test data, however, as Hambleton and Jones (1993) note, “models always provide incomplete representations of the test data to which they are fit” (p. 39). Therefore, models are always in some manner inaccurate, yet help us simplify the complexity of reality. Measurement models that fit the data reasonably well are useful for representing a construct, and the process of applying measurement models to data is valuable in the scale development process because it can highlight problems with our measurements or even cause us to reconsider our theory of how the construct should be represented.

Evaluating measurement models is particularly important for the study of temperament, as temperament is measured from infancy through adolescence and into adulthood with different scales using items tailored to developmental age. Furthermore, our understanding of the structure of temperament has been informed by measurement models depicting heterotypic change in higher-order factor structure (Putnam, Gartstein, & Rothbart, 2008). For example, perceptual sensitivity appears as a component of effortful control on the CBQ at ages 3-7, but the corresponding scale loads on surgency in infancy measured with the IBQ-R (Gartstein & Rothbart, 2003) and on negative affectivity in early childhood measured with the ECBQ (Putnam, Gartstein, & Rothbart, 2006). While the factor of effortful control may demonstrate heterotypic changes over time (Putnam, Rothbart, & Gartstein, 2008) individual scales should be less susceptible to changes in composition. Perceptual sensitivity should still measure the same underlying construct—sensitivity to minor changes in the environment—even though its measurement may need to be adapted with age. For instance, it would be inappropriate to ask parents how often their infant comments if someone has an unusual voice. Temperament is relatively stable (Rothbart & Bates, 2006), and therefore specific traits should not change dramatically over time but how those qualities are measured and relate to other qualities may.

Thus, I would argue that demonstrating longitudinally stable, valid, and reliable item-level measurement models of temperament constructs is of utmost importance in understanding how temperament changes over time. If a single construct is measured adequately and consistently over time, only then can we begin to examine how broader factors of temperament develop. In this way, the measurement of temperament may further inform the theory of temperament.

To truly establish the longitudinal consistency of a measurement model requires longitudinal data. Yet, during the construction and evaluation phases of scale development, comparing measurement models for two independent samples that differ in age is a more appropriate alternative, because time and resources are not dedicated to following a sample over several years when psychometric properties of the scale have not yet been tested. There are several steps that can be taken that may provide evidence of longitudinal consistency using two samples. First, data must be collected from younger and older samples roughly equivalent in terms of demographics (race/ethnicity, SES). Next items should demonstrate adequate psychometric properties in both samples (e.g., low incidence of missing or N/A responses, reasonable variability of responses). Measurement invariance should then be tested—performed in a confirmatory factor analysis (CFA) context—which establishes the equivalence of measurement models and consistency with which a latent construct is measured across age groups (Brown, 2006). A more thorough treatment of measurement invariance testing is provided in Study 1 below. Finally, all items in the final measurement model should be salient to the measurement of the latent construct, and thus load on the construct at an acceptable level (e.g., $\lambda < .30$; Brown, 2006).

We can apply these guidelines to the CBQ EC scales quite easily. As the primary purpose of evaluating the EC scales is to investigate potential measurement invariance across age groups, CBQ data for young children will be compared to that for older children. Two samples will be needed, each large enough to sufficiently estimate CFA parameters. Next response frequency will be examined, followed by testing for factor structure and measurement invariance. Recommendations for revising items, scales, and the response scale will be based on these findings.

# CHAPTER 2

# STUDY 1: EVALUATING THE CBQ EFFORTFUL CONTROL SCALES

## Purpose

The purpose of this study is to evaluate the CBQ EC scales to identify items that have desirable psychometric attributes for measuring EC constructs adequately and consistently between ages three and seven.

## Method

### Participants

The present study uses archival data collected from multiple sources (Table 2.1). Most data were those used in the CBQ-SF and CBQ-VSF validation study (Putnam & Rothbart, 2006). Additional data were provided by Dr. John Bates at the University of Indiana, Dr. Victoria Molfese and Dr. Kathleen Moritz Rudasill at the University of Nebraska-Lincoln, Dr. Samuel Putnam at Bowdoin College, and Dr. Leslie Leve at the University of Oregon. Data from all sources represent CBQ or CBQ-SF data on children ($N$ = 2,469) age 21 months to 101 months. Because the CBQ is intended for children ages three to seven years, data for children 35 months and younger and 96 months and older were omitted from analyses. In addition, the purpose of the present study is to compare the performance of CBQ scales and items between younger and older children, so the decision was made to select two subsamples from this dataset for comparison: children ages 3-4 (the “younger” sample) and children ages 6-7 (the “older” sample). Children age five (60 months to 71 months), were omitted from analyses because (a) measurement invariance applying to item data between 3-4 year olds and 6-7 year olds should also

Table 2.1

*Sample descriptives for CBQ Evaluation by age range and source.*

| | | | Child Gender | | |
|---|---|---|---|---|---|
| Ages 3-4 | Form | Total | Girl | Boy | % Girl |
| Bates 1 | CBQ-SF | 89 | 40 | 49 | 44.9% |
| Bates 2 | CBQ-SF | 93 | 40 | 53 | 43.0% |
| Carlson1* | CBQ | 229 | 123 | 106 | 53.7% |
| Carlson2* | CBQ-SF | 100 | 47 | 53 | 47.0% |
| Cole* | CBQ-SF | 289 | 136 | 153 | 47.1% |
| Fagot & Leve (1998)* | CBQ | 8 | 6 | 2 | 75.0% |
| Kochanska (1997)* | CBQ | 99 | 48 | 51 | 48.5% |
| Kochanska et al. (1994)* | CBQ | 91 | 41 | 50 | 45.1% |
| Molfese & Rudasill | CBQ-SF | 104 | 48 | 56 | 46.2% |
| Putnam | CBQ | 187 | 85 | 102 | 45.5% |
| Rothbart* | CBQ | 57 | 28 | 29 | 49.1% |
| Victor* | CBQ-SF | 62 | 30 | 32 | 48.4% |
| **Total - Ages 3-4** | | **1408** | **672** | **736** | **47.7%** |
| Ages 6-7 | | | | | |
| Fagot & Leve* | CBQ | 30 | 12 | 18 | 40.0% |
| Fisher* | CBQ | 113 | 50 | 63 | 44.2% |
| Gunnar* | CBQ | 59 | 30 | 29 | 50.8% |
| LeMare* | CBQ | 54 | 21 | 33 | 38.9% |
| Leve | CBQ-SF alt | 331 | 144 | 187 | 43.5% |
| **Total - Ages 6-7** | | **587** | **257** | **330** | **43.8%** |
| **Grand Total** | | **1995** | **929** | **1066** | **46.6%** |

CBQ = Children's Behavior Questionnaire. CBQ-SF = Children's Behavior Questionnaire Short Form. CBQ-SF alt = CBQ-SF with extra Inhibitory Control items.
* Sample used in Putnam & Rothbart (2006)

apply to 5 year olds, (b) the alternative approach of arbitrarily splitting the sample by age presents the dilemma of separating some children who differ in age by only a few months, and treating them as categorically distinct, and (c) children age 5 are often beginning school entry, and by focusing on younger and older children we may avoid parent-ratings of children who are experiencing new and complex social and academic demands. I therefore judged having two samples ages 3-4 and 6-7 would allow for clearer developmental distinction between groups and more meaningful interpretations of group differences that may be found. From these selection criteria, a large final sample (*N* = 1,995) was retained, with the younger sample (*N* = 1,408, 672 girls) considerably larger than the older sample (*N*=587, 257 girls). Although an unbalanced design may affect multi-group Type I error rate with smaller samples (i.e., N < 100), the samples in the present study are considerably large (i.e., N > 200) and provide enough power for this to not be of concern (Brown, 2006).

Data contributed from most sources used by Putnam and Rothbart (2006) did not include ethnicity or socioeconomic status, so it was not possible to statistically compare the younger sample to the older sample on these demographics. However, Putnam and Rothbart did provide brief descriptions of these samples, and most are reported to be primarily White, middle-to-upper class families. The sample provided by Dr. Robert Cole's laboratory (N = 229, 16.3% of younger sample data) was identified as being particularly diverse, and reported to be about 34% White and 66% Black or other ethnicity, with approximately half of the sample lower income. Therefore, it is possible the younger sample contained proportionally fewer White and middle-to-upper families than the older sample.

**Measures**

Archival data from the CBQ Standard Form (Rothbart, Ahadi, Hershey, & Fisher, 2001) and the CBQ-SF (Putnam & Rothbart, 2006) were used in the present study. Of the CBQ data (Ages 3-4 *N* = 671; Ages 6-7 *N* = 256; Total *N* = 927) and CBQ-SF data (Ages 3-4 *N* = 737; Ages 6-7 *N* = 331; Total *N* = 1,068), a majority (69.2%) did not have information on relation of the reporter to the child, although for data with such information, mother-report was most common (29.6%), with some father (1.2%) and alternative caregiver report (0.2%). The CBQ EC scales are comprised of 52 items: Attention Shifting (5 items), Attentional Focusing (9 items), Inhibitory Control (13 items), Low Intensity Pleasure (13 items), and Perceptual Sensitivity (12 items). The CBQ-SF EC scales are comprised of 26 items: Attentional Focusing (6 items), Inhibitory Control (6 items), Low Intensity Pleasure (8 items), and Perceptual Sensitivity (6 items). CBQ-SF data from Leslie Leve are parent report of adopted children, and included a full (13 item) inhibitory control scale. The CBQ and CBQ-SF use a 7-point rating scale, with the following anchors: 1 = “extremely untrue of your child,” 2 = “quite untrue of your child,” 3 = “slightly untrue of your child,” 4 = “neither true nor false of your child,” 5 = “slightly true of your child,” 6 = “quite true of your child,” 7 = “extremely true of your child.”

**Analysis**

First, scale descriptives and scale score mean differences were calculated for both age groups and each form (standard vs. short). Independent samples t-tests were conducted for each form of each scale to examine whether scale scores were significantly different between the older and younger age groups. Given the developmental nature of effortful control, it was expected that older children should generally score higher than younger children. Cohen’s *d* values were calculated as effect sizes.

Next, the rating scales were evaluated by examining response option use. For each group, the response option use for every item was calculated as a percentage: the number respondents who selected the option divided by the total item responses. These were then aggregated in two ways. First, for every scale I calculated the percent of each option used by less than 5% of respondents. For a more restrictive analysis, I also calculated the percent of each option used by less than 1% of respondents. If multiple scales show low response frequency for certain response options, this may suggest that the rating scale could be modified, or that more items are needed to better cover the scale range. Six decision rules were then set for selecting CBQ items to retain for revised EC scales (Table 2.2). When constructing the CBQ-SF from the standard scale, Putnam and Rothbart (2006) retained only items for which at least 80% of respondents selected a response on the 1-7 rating scale (rather than marking "n/a"). Decision Rule 1 in the present study employs this criterion.

It has been suggested that items with multiple categories with negligible use (< 1% of responses) should be considered for revision (Penfield, 2013). This is because items with several infrequently used response options have less variability and therefore are less able to differentiate between individuals. This criterion is reflected in Decision Rule 2, with items flagged if they had three or more response categories with negligible use in either sample.

Decision Rules 3, 4, and 5 involve testing for measurement invariance. Measurement invariance is important to establish when an instrument is administered to a heterogeneous population (Brown, 2006). In the current context, the CBQ is intended for use with children ages three to seven, and for reasons identified in the previous section, it is likely that items intended to measure effortful control may not operate in the same manner for younger children (i.e., 3-4 year olds) as for somewhat older children (i.e., 6-7 year olds). If an item is shown to be noninvariant, then it does not measure the underlying construct the same way across the groups for which invariance was tested, and is thus

Table 2.2

*Decision rules for CBQ Evaluation.*

| | Decision Rule | Source |
|---|---|---|
| 1 | Missing 20% + of data for either group | Putnam & Rothbart (2006) |
| 2 | 3+ response categories with negligible (<1%) data for either group | Penfield (2013) |
| 3* | Configural invariance across group | Brown (2006) |
| 4* | Presence of metric (weak) noninvariance | Brown (2006); Clark et al., (2016) |
| 5* | Presence of scalar (strong) noninvariance | Brown (2006); Clark et al., (2016) |
| 6 | Factor loadings < .30 | Brown (2006) |

* Decision rule applies to context of measurement invariance testing.

biased (Brown, 2006). Invariance testing can be conducted using confirmatory factor analysis (CFA) with the structures and parameters of the measurement model compared across two or more groups. In the current study, structural models are compared for younger (ages 3-4) and older (ages 6-7) groups.

Testing for measurement invariance involves several sequential steps. First, the factor structures for each group are tested for equivalence, known as *configural invariance*. Next, I tested for *metric invariance* by constraining factor loadings to be equal across groups, such that the loadings for Item 1 are equal, loadings for Item 2 are equal, and so on. Finally, I tested for *scalar invariance* by constraining the item intercepts across groups. I will now discuss the specific processes involved in these steps in detail.

When testing for configural invariance, measurement models are estimated separately for each group via CFA, with the expectation that all indicators should have significant factor loadings for both groups. Both models should fit according to goodness-of-fit indices such as Root Mean Square Error of Approximation (RMSEA) and Comparative Fit Index (CFI) (Bentler, 1990; Brown, 2006; Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008). Chi-square tests of fit are also commonly used, but are very sensitive to small degrees of misfit with large sample (Brown, 2006), and therefore will not be used in the present study. In this study, the factor structures are identified by initially standardizing the factor scales to a mean of 0.0 and a variance of 1.0 (although factor means are freed during testing for scalar invariance). By fixing the factor scales, all item loadings can be freely estimated. If an indicator had a non-significant loading (loading is not significantly different from 0), this suggests the item does not sufficiently relate to the latent construct and should be omitted from subsequent invariance analyses. The convention for CFA analyses is that items with factor loadings less than .30 are also considered not meaningfully related to the factor and therefore may not be useful to the model (Brown, 2006); this guideline is incorporated into Decision Rule 6. However, if an item had a non-significant factor loading when testing the initial factor structure in either group, that item was flagged for violating Decision Rule 3, *configural invariance*, and omitted from subsequent invariance analyses (Brown, 2006).

RMSEA values less than .08 are considered "acceptable" and values less than .05 indicate "close" fit, whereas CFI values greater than .90 are considered acceptable (Bentler, 1990; Brown, 2006; Cheung & Rensvold, 2002). If a model demonstrates poor fit, modification indices may be consulted to identify additional residual covariance parameters that may be added to improve model-data fit. To ensure the structure of both groups are identical, it is important to take adjustments made for one group and apply them to the other group. Once the factor structures from both groups are the same and demonstrate acceptable fit, I formally tested for configural invariance by simultaneously estimating the factor structures for both groups in a single model and confirming the presence of satisfactory fit. This model is known as the *configural model*, and is the reference against which the nested *metric models* are compared.

Next, I tested for metric invariance by estimating the configural model with factor loadings constrained to be equal across groups. If metric *noninvariance* is found for a certain item, this indicates that the relationship between the item and the factor changes at different rates across groups. For instance, an item with a high factor loading for group A and a lower factor loading for group B is more closely related to the latent factor for group A and therefore the item changes more drastically from fluctuations in the latent factor in group A than for group B. Because this *metric model* is nested within the configural model (i.e., structurally equivalent, but with fewer freely estimated parameters), I compared how much more poorly the metric model fit the data compared to the configural model using criteria employed in another study in which CBQ was evaluated for measurement invariance by gender (Clark et al., 2016) and based on guidelines of Meade et al., (2008). First, a loglikelihood difference (Δ-2LL) test was evaluated. The difference between the model loglikelihoods*(-2) follows a chi-square ($\chi^2$) distribution, where the degrees of freedom are the difference in freely estimated model parameters (Brown, 2006). A non-significant $\chi^2$-statistic ($p > .05$) indicates that the nested (metric) model fits *not significantly worse than* the reference (configural) model. Second, the difference in comparative fit indices (ΔCFI) was evaluated. Meade et al. (2008) suggested that ΔCFI < .002 also indicate the nested model fits comparably to the fit of the reference model. Third, the difference in Root Mean Square Error of Approximation (ΔRMSEA) values was evaluated. Meade et al. (2008) suggested that ΔRMSEA < .015 indicate a nested model that fits as well as the reference model. In line with Clark and colleagues (2016), for a model to be accepted at each stage, at least two out of three of these criteria must be met. If the nested model showed poor fit, the item demonstrating the greatest misfit via modification indices was freed in a subsequent model (i.e., *metric model #2*), which then was compared to the reference configural model using the 2-out-of-3 fit criteria. If the model still fit significantly worse than the configural model, another loading was freed in a subsequent model, compared to the configural model, and so on until a metric model was found that was not significantly worse than the configural model. This *final metric model* became the new base against which the *scalar models* were compared.

Scalar invariance was tested by estimating the final metric model but constraining the item intercepts to be equal across groups, while at the same time allowing the factor mean of one group to be freely estimated. Only the intercepts of items which demonstrated metric invariance were constrained to be equal; the intercepts of metric noninvariant items remained free. Scalar invariance is required so that mean differences in factor scores can be meaningfully compared across groups, as observed item mean differences are attributable solely to the latent factor (Millsap & Olivera-Aguilar, 2012). If scalar noninvariance is found for an item, then a difference in the observed item means between two groups cannot be solely attributable to mean differences in the latent factors. Scalar models were compared to the reference final metric model using the same 2-out-of-3 fit criteria outlined above. If the scalar model fit significantly worse than the reference model, the constrained intercept was freed for the item demonstrating greatest misfit. The new scalar model was then compared to the final reference model, and this process continued until a scalar model was found to be comparable to the final metric model.

## Results

Scale descriptive statistics and group differences are presented in Table 2.3. Low Intensity Pleasure had the highest average score across age groups and CBQ form (5.65 to 6.04), whereas Attention Shifting had the lowest average score (3.95 to 3.98). Group mean differences were calculated via independent samples t-test, with Cohen's *d* for effect size. Children ages 6-7 were rated significantly higher than children ages 3-4 on both the standard and short forms of the Attentional Focusing ($t$ = 6.08 and $t$ = 3.67, *ps* < .001) and Inhibitory Control ($t$ = 7.60 and $t$ = 6.50, *ps* < .001) scales, yet were rated lower on the Low Intensity Pleasure subscale on the short form ($t$ = -9.05, $p$ < .001). These effect sizes were small to medium in strength ($|d|$ = .18 to .54). To further probe age differences in Low Intensity Pleasure, item-level descriptives and group differences are presented in Table 2.5. Six items showed significantly lower means ($p$ < .01) for children age 6-7 than for children age 3-4.

Response frequency analyses are presented in Table 2.5 (<5% use) and Table 2.6 (<1% use). The percentages listed in these tables indicate the percent of items within each scale for which a given response option has low (<5%) or negligible (<1%) use, respectively, broken down by CBQ form and age group. All data were used for the Short Form analyses (both the standard and short forms contain all CBQ-SF items), and only data from sources using the standard form CBQ were used for calculating response frequency for the standard form. Table 2.5 and Table 2.6 incorporate color scales, so that response options with darker fill colors indicate less frequent use. For instance, examine the Attentional Focusing scale in Table 2.5. For the younger group (ages 3-4), the second response option, corresponding to the anchor “quite untrue of your child” has low use in 22.2% of items on this scale. Comparatively, this response option was used less frequently with the older age group, with 55.6% of items on the scale having low use of this option. From reviewing Table 2.5, it seems that the lowest response category, corresponding to the anchor “extremely untrue of your child” (and “extremely true of your child” for reverse-worded items) is generally infrequently used, and this is supported by examining the corresponding option response use in Table 2.6. Furthermore, the lower four response options are seldom used with the Low Intensity Pleasure scale, and multiple response options show moderate amounts of negligible use.

Table 2.3

*Descriptive statistics and group differences for CBQ and CBQ-SF EC scales.*

| | | | Ages 3-4 Years | | | Ages 6-7 Years | | | Group Differences | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Scale | Form | # Items | N | Mean | SD | N | Mean | SD | *t* | *p* | *d* |
| Attentional Focusing | CBQ | 9 | 671 | 4.60 | 0.86 | 256 | 4.98 | 0.79 | 6.08 | < .001 | 0.46 |
| | CBQ-SF | 6 | 1408 | 4.77 | 0.98 | 587 | 4.95 | 1.03 | 3.67 | < .001 | 0.18 |
| Attention Shifting | CBQ | 5 | 671 | 3.95 | 0.90 | 256 | 3.98 | 1.02 | 0.49 | *ns* | 0.03 |
| Inhibitory Control | CBQ | 13 | 671 | 4.57 | 0.77 | 256 | 5.01 | 0.86 | 7.60 | < .001 | 0.54 |
| | CBQ-SF | 6 | 1408 | 4.51 | 0.93 | 587 | 4.81 | 0.96 | 6.50 | < .001 | 0.32 |
| Low Intensity Pleasure | CBQ | 13 | 671 | 5.70 | 0.59 | 256 | 5.65 | 0.61 | -1.16 | *ns* | -0.08 |
| | CBQ-SF | 8 | 1408 | 6.04 | 0.61 | 587 | 5.76 | 0.65 | -9.05 | < .001 | -0.44 |
| Perceptual Sensitivity | CBQ | 12 | 671 | 4.99 | 0.79 | 256 | 5.06 | 0.78 | 1.10 | *ns* | 0.08 |
| | CBQ-SF | 6 | 1408 | 5.39 | 0.90 | 587 | 5.39 | 0.88 | -0.03 | *ns* | 0.00 |

Table 2.4

*Descriptive statistics and group differences for CBQ Low Intensity Pleasure scale.*

| | | Ages 3-4 Years | | | Ages 6-7 Years | | | Group Differences | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CBQ# | Item | N | Mean | SD | N | Mean | SD | *t* | *p* | *d* |
| 12r | Rarely enjoys just being talked to. | 669 | 5.71 | 1.49 | 255 | 5.79 | 1.39 | 0.78 | *ns* | .058 |
| 36 | Enjoys just sitting quietly in the sunshine. | 660 | 3.44 | 1.62 | 254 | 3.56 | 1.64 | 1.04 | *ns* | .077 |
| 54 | Enjoys taking warm baths. | 1399 | 6.26 | 1.06 | 582 | 5.88 | 1.14 | -7.07 | <.001 | -.344 |
| 66r | Doesn't enjoy being read to very much. | 671 | 6.55 | 0.92 | 256 | 6.52 | 1.02 | -0.47 | *ns* | -.034 |
| 76 | Enjoys "snuggling up" next to a parent. | 1404 | 6.39 | 0.92 | 585 | 6.31 | 1.06 | -1.71 | *ns* | -.082 |
| 86r | Doesn't care much for quiet games. | 667 | 4.77 | 1.50 | 255 | 5.32 | 1.25 | 5.19 | <.001 | .397 |
| 111r | Isn't interested in watching quiet TV shows, such as "Mister Rogers." | 640 | 5.15 | 1.64 | 248 | 4.82 | 1.85 | -2.60 | <.01 | -.189 |
| 113 | Enjoys just being talked to. | 1390 | 6.02 | 1.12 | 584 | 5.91 | 1.07 | -2.11 | *ns* | -.105 |
| 133 | Enjoys looking at picture books. | 1406 | 6.33 | 0.86 | 585 | 5.99 | 0.99 | -7.78 | <.001 | -.371 |
| 146 | Likes being sung to. | 1392 | 5.80 | 1.31 | 575 | 5.39 | 1.32 | -6.32 | <.001 | -.313 |
| 151 | Likes the sound of words, as in nursery rhymes. | 1390 | 5.97 | 0.97 | 584 | 5.70 | 1.00 | -5.78 | <.001 | -.283 |
| 164 | Enjoys gentle rhythmic activities, such as rocking or swaying. | 1396 | 5.35 | 1.33 | 577 | 4.82 | 1.53 | -7.68 | <.001 | -.369 |
| 174 | Enjoys sitting on parent's lap. | 1404 | 6.19 | 1.01 | 584 | 6.08 | 1.03 | -2.33 | *ns* | -.114 |

Table 2.5

*Percent CBQ and CBQ-SF EC items with low response option use (<5%).*

| Scale | Age Group | N | Response Option* 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| Combined EC Scales | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 86.5% | 42.3% | 21.2% | 46.2% | 1.9% | 0.0% | 32.7% |
| (52 items) | 6-7 yrs | 256 | 84.6% | 51.9% | 25.0% | 40.4% | 1.9% | 0.0% | 21.2% |
| Short Form | 3-4 yrs | 1406 | 92.3% | 53.8% | 34.6% | 30.8% | 0.0% | 0.0% | 11.5% |
| (26 items) | 6-7 yrs | 587 | 96.2% | 61.5% | 34.6% | 38.5% | 0.0% | 0.0% | 3.8% |
| Attentional Focusing | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 88.9% | 22.2% | 0.0% | 33.3% | 0.0% | 0.0% | 44.4% |
| (9 items) | 6-7 yrs | 256 | 88.9% | 55.6% | 11.1% | 44.4% | 0.0% | 0.0% | 11.1% |
| Short Form | 3-4 yrs | 1406 | 100.0% | 16.7% | 0.0% | 0.0% | 0.0% | 0.0% | 16.7% |
| (6 items) | 6-7 yrs | 587 | 100.0% | 33.3% | 0.0% | 50.0% | 0.0% | 0.0% | 0.0% |
| Attention Shifting | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 60.0% | 0.0% | 0.0% | 40.0% | 0.0% | 0.0% | 80.0% |
| (5 items) | 6-7 yrs | 256 | 80.0% | 20.0% | 0.0% | 40.0% | 0.0% | 0.0% | 100.0% |
| Inhibitory Control | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 92.3% | 46.2% | 0.0% | 46.2% | 0.0% | 0.0% | 46.2% |
| (13 items) | 6-7 yrs | 587 | 92.3% | 30.8% | 7.7% | 38.5% | 0.0% | 0.0% | 30.8% |
| Short Form | 3-4 yrs | 1406 | 66.7% | 33.3% | 0.0% | 16.7% | 0.0% | 0.0% | 33.3% |
| (6 items) | 6-7 yrs | 587 | 83.3% | 50.0% | 0.0% | 50.0% | 0.0% | 0.0% | 16.7% |

Table 2.5, continued

*Percent CBQ and CBQ-SF EC items with low response option use (<5%).*

| Scale | Age Group | N | Response Option* 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| Low Intensity Pleasure | | | | | | | | | |
| Standard Form (13 items) | 3-4 yrs | 671 | 92.3% | 76.9% | 61.5% | 76.9% | 7.7% | 0.0% | 7.7% |
| | 6-7 yrs | 256 | 84.6% | 84.6% | 69.2% | 61.5% | 7.7% | 0.0% | 7.7% |
| Short Form (8 items) | 3-4 yrs | 1406 | 100.0% | 100.0% | 87.5% | 75.0% | 0.0% | 0.0% | 0.0% |
| | 6-7 yrs | 587 | 100.0% | 87.5% | 75.0% | 37.5% | 0.0% | 0.0% | 0.0% |
| Perceptual Sensitivity | | | | | | | | | |
| Standard Form (12 items) | 3-4 yrs | 671 | 83.3% | 33.3% | 25.0% | 25.0% | 0.0% | 0.0% | 16.7% |
| | 6-7 yrs | 256 | 83.3% | 50.0% | 16.7% | 16.7% | 0.0% | 0.0% | 8.3% |
| Short Form (6 items) | 3-4 yrs | 1406 | 100.0% | 50.0% | 33.3% | 16.7% | 0.0% | 0.0% | 0.0% |
| | 6-7 yrs | 587 | 100.0% | 66.7% | 50.0% | 16.7% | 0.0% | 0.0% | 0.0% |

*Accounting for reverse-scored items.

Color Scale: 100% | > 75% | > 50% | > 25%

Table 2.6

*Percent CBQ and CBQ-SF EC items with negligible response option use (<1%).*

| | | | Response Option* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Scale | Group | N | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Combined EC Scales | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 44.2% | 13.5% | 3.8% | 7.7% | 0.0% | 0.0% | 3.8% |
| (52 items) | 6-7 yrs | 256 | 42.3% | 17.3% | 0.0% | 1.9% | 0.0% | 0.0% | 1.9% |
| Short Form | 3-4 yrs | 1406 | 42.3% | 15.4% | 0.0% | 3.8% | 0.0% | 0.0% | 0.0% |
| (26 items) | 6-7 yrs | 587 | 50.0% | 7.7% | 0.0% | 3.8% | 0.0% | 0.0% | 0.0% |
| Attentional Focusing | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 55.6% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| (9 items) | 6-7 yrs | 256 | 33.3% | 11.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Short Form | 3-4 yrs | 1406 | 16.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| (6 items) | 6-7 yrs | 587 | 16.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Attention Shifting | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| (5 items) | 6-7 yrs | 256 | 20.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Inhibitory Control | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 38.5% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 15.4% |
| (13 items) | 6-7 yrs | 587 | 30.8% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 7.7% |
| Short Form | 3-4 yrs | 1406 | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| (6 items) | 6-7 yrs | 587 | 33.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |

Table 2.6, continued

*Percent CBQ and CBQ-SF EC items with negligible response option use (<1%).*

| | | | Response Option* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Scale | Group | N | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Low Intensity Pleasure | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 61.5% | 53.8% | 15.4% | 30.8% | 0.0% | 0.0% | 0.0% |
| (13 items) | 6-7 yrs | 256 | 69.2% | 46.2% | 0.0% | 7.7% | 0.0% | 0.0% | 0.0% |
| Short Form | 3-4 yrs | 1406 | 62.5% | 50.0% | 0.0% | 12.5% | 0.0% | 0.0% | 0.0% |
| (8 items) | 6-7 yrs | 587 | 87.5% | 25.0% | 0.0% | 12.5% | 0.0% | 0.0% | 0.0% |
| Perceptual Sensitivity | | | | | | | | | |
| Standard Form | 3-4 yrs | 671 | 25.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| (12 items) | 6-7 yrs | 256 | 41.7% | 16.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Short Form | 3-4 yrs | 1406 | 33.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| (6 items) | 6-7 yrs | 587 | 50.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |

*Accounting for reverse-scored items.

Color Scale: > 75% > 50% > 25%

Decision rules are summarized in Table 2.1, and results are presented in Table 2.7. Based on Decision Rule 1 one item from the inhibitory control scale was flagged for having considerable “N/A” responses (CBQ 20: *“Is good at games like ‘Simon Says,’ ‘Mother, May I?’ and ‘Red Light, Green Light’”*). A sizeable portion of both groups marked “n/a” on this item: 37.6% of responses for the 3-4 year old sample and 20.3% of the 6-7 year old sample. Two items were flagged based on Decision Rule 2, both appearing on the Low Intensity Pleasure subscale (CBQ 66: *“Doesn't enjoy being read to very much”*; CBQ 133: *“Enjoys looking at picture books”*). For these items, the lower two options (“extremely untrue your child” and “quite untrue of your child”) and middle option (“neither true nor false of your child”) exhibited negligible use (<1% of responses) in the 3-4 year old sample.

Five items demonstrated configural noninvariance (Decision Rule 3) and were flagged: two from the Attentional Focusing scale (CBQ 160. *“Has difficulty leaving a project s/he has begun”*; CBQ 186. *“Sometimes becomes absorbed in a picture book and looks at it for a long time”*); two from Low Intensity Pleasure (CBQ 54. *“Enjoys taking warm baths”* and CBQ 111r. *“Isn't interested in watching quiet TV shows, such as ‘Mister Rogers’”*); and one from Perceptual Sensitivity (CBQ 142r. *“Doesn't usually react to different textures of food”*). All items had non-significant loadings for the older group except for CBQ 54, which had a non-significant loading for the young group. CBQ 111r had non-significant loadings for both groups. These items were dropped from subsequent analyses.

Table 2.7

*Results of CBQ EC item evaluation by decision rule.*

| Scale | CBQ# | Item Stem |
|---|---|---|
| Item Flagged by Decision Rule 1 (>20% n/a responses) | | |
| Inhi | 20 | Is good at games like "Simon Says," "Mother, May I?" and "Red Light, Green Light." |
| Items Flagged by Decision Rule 2 (3+ response categories with negligible (<1%) data for either group) | | |
| Lopl | 66 | Doesn't enjoy being read to very much. |
| Lopl | 133 | Enjoys looking at picture books. |
| Items Flagged by Decision Rule 3 (Configural noninvariance) | | |
| AttF | 160 | Has difficulty leaving a project s/he has begun. |
| AttF | 186 | Sometimes becomes absorbed in a picture book and looks at it for a long time. |
| Lopl | 54 | Enjoys taking warm baths. |
| Lopl | 111 | Isn't interested in watching quiet TV shows, such as "Mister Rogers." |
| Perc | 142 | Doesn't usually react to different textures of food. |
| Items Flagged by Decision Rule 4 (Metric noninvariance) | | |
| AttF | 38 | When practicing an activity, has a hard time keeping her/his mind on it. |
| AttS | 184 | Sometimes doesn't seem to hear me when I talk to her/him. |
| Inhi | 4 | Can lower his/her voice when asked to do so. |
| Inhi | 185 | Is usually able to resist temptation when told s/he is not supposed to do something. |
| Perc | 28 | Usually doesn't comment on changes in parents' appearance. |
| Perc | 98 | Is quickly aware of some new item in the living room. |

Table 2.7, continued

*Results of CBQ EC item evaluation by decision rule.*

| Scale | CBQ# | Item Stem |
|---|---|---|
| Items Flagged by Decision Rule 5: (Scalar noninvariance) | | |
| AttF | 125 | When drawing or coloring in a book, shows strong concentration. |
| AttF | 195 | Has a hard time concentrating on an activity when there are distracting noises. |
| AttS | 29 | Can easily shift from one activity to another. |
| Inhi | 63 | Prepares for trips and outings by planning things s/he will need. |
| Inhi | 93 | Has difficulty waiting in line for something. |
| Inhi | 108 | Has trouble sitting still when s/he is told to (at movies, church, etc.). |
| Inhi | 116 | Is able to resist laughing or smiling when it isn't appropriate. |
| Inhi | 162 | Is not very careful and cautious in crossing streets. |
| Lopl | 86 | Doesn't care much for quiet games. |
| Lopl | 146 | Likes being sung to. |
| Lopl | 151 | Likes the sound of words, as in nursery rhymes. |
| Lopl | 164 | Enjoys gentle rhythmic activities, such as rocking or swaying. |
| Perc | 9 | Notices the smoothness or roughness of objects s/he touches. |
| Perc | 52 | Seems to listen to even quiet sounds. |
| Perc | 65 | Comments when a parent has changed his/her appearance. |
| Perc | 105 | Usually comments if someone has an unusual voice. |
| Perc | 122 | Does not seem to notice parents' facial expressions. |
| Perc | 154 | Notices even little specks of dirt on objects. |

Table 2.7, continued

*Results of CBQ EC item evaluation by decision rule.*

| Scale | CBQ# | Item Stem |
|---|---|---|
| Items Flagged by Decision Rule 6 (Factor Loading < .30) | | |
| Lopl | 12 | Rarely enjoys just being talked to. |
| Perc | 84 | Doesn't usually comment on people's facial features, such as size of nose or mouth. |
| Remaining Items | | |
| AttF | 16 | When picking up toys or other jobs, usually keeps at the task until it's done. |
| AttF | 47 | Will move from one task to another without completing any of them. |
| AttF | 144 | When building or putting something together, becomes very involved in what s/he s doing, and works for long periods. |
| AttF | 171 | Is easily distracted when listening to a story. |
| AttS | 6 | Is hard to get her/his attention when s/he is concentrating on something. |
| AttS | 95 | Has a lot of trouble stopping an activity when called to do something else. |
| AttS | 180 | Has an easy time leaving play to come to dinner. |
| Inhi | 32 | Has a hard time following instructions. |
| Inhi | 75 | Can wait before entering into new activities if s/he is asked to. |
| Inhi | 136 | Is good at following instructions. |
| Inhi | 147 | Approaches places s/he has been told are dangerous slowly and cautiously. |
| Inhi | 168 | Can easily stop an activity when s/he is told "no." |
| Lopl | 36 | Enjoys just sitting quietly in the sunshine. |
| Lopl | 76 | Enjoys "snuggling up" next to a parent. |
| Lopl | 113 | Enjoys just being talked to. |
| Lopl | 174 | Enjoys sitting on parent's lap. |
| Perc | 31 | Notices it when parents are wearing new clothing. |
| Perc | 170 | Doesn't usually notice odors, such as perfume, smoke, cooking, etc. |

Rev = Item is reversed when computing scale. AttF = Attentional Focusing; AttS = Attention Shifting; Inhi = Inhibitory Control; Lopl = Low intensity pleasure; Perc = Perceptual Sensitivity.

Six items demonstrated metric noninvariance (Decision Rule 4) and were flagged: one from Attentional Focusing (CBQ 38r. *"When practicing an activity, has a hard time keeping her/his mind on it"*); one from Attention Shifting (CBQ 184r. *"Sometimes doesn't seem to hear me when I talk to her/him"*); two from Inhibitory Control (CBQ 4. *"Can lower his/her voice when asked to do so"* and CBQ 185. *"Is usually able to resist temptation when told s/he is not supposed to do something"*); and two from Perceptual Sensitivity (CBQ 28r. *"Usually doesn't comment on changes in parents' appearance"* and CBQ 98. *"Is quickly aware of some new item in the living room"*). All of these items demonstrated stronger loadings in the older group except CBQ 185 and CBQ 28r, which were stronger in the younger group.

Eighteen items demonstrated scalar noninvariance (Decision Rule 5) and were flagged: two from Attentional Focusing; one from Attention Shifting; five from Inhibitory Control; four from Low Intensity Pleasure; and six from Perceptual Sensitivity. Factor loadings were then evaluated using the final scalar model for each scale. Two items had unacceptable factor loadings (<.30): one from Low Intensity Pleasure (CBQ 12r. *"Rarely enjoys just being talked to"*); one from Perceptual Sensitivity (CBQ 84r. *"Doesn't usually comment on people's facial features, such as size of nose or mouth"*).

Eighteen items remained after employing Decision Rules 1 through 5: four items from Attentional Focusing; three items from Attention Shifting; five items from Inhibitory Control; four items from Low Intensity Pleasure; and two items from Perceptual Sensitivity. However, two remaining pairs of items required covarying residuals to achieve adequate model fit in the initial measurement models (prior to estimating the configural model). These covariances were retained throughout subsequent models and pose an issue related to the measurement of the underlying construct. Confirmatory factor analysis (CFA) estimates the relationships between observed items as a function of a latent factor. Ideally, the latent factor should be the only commonality between items, meaning that after accounting for the factor, no remaining covariance should exist among items' residuals. A relation that does exist between two items after accounting for the factor, in the form of covaried residual term, indicates that something explains the relation between the items *over and above* that explained by the factor (Brown, 2006). This may arise for a variety of reasons, such as similar wording, content overlap, or if the items are the reverse of one another. This may be undesirable because it can bias CFA and structural equation model (SEM) parameter estimates (Tomarken & Waller, 2001), and may over-represent the items and their content when measuring the construct (Gerbing & Anderson, 1984). One set of remaining items with covaried residuals are from the Inhibitory Control scale (CBQ 32r. *"Has a hard time following instructions"* and CBQ 136. *"Is good at following instructions"*), and clearly cover identical content only with reverse wording. The other set of items are from the Low Intensity Pleasure scale (CBQ 76. *"Enjoys 'snuggling up' next to a parent"* and CBQ 174. *"Enjoys sitting on parent's lap"*), which have very similar content: enjoying physical contact with parents.

## Discussion

This study demonstrates that numerous flaws exist with existing CBQ EC scales and items. First, inconsistencies in mean scale scores were found, primarily for the Low Intensity Pleasure scale. Second, response frequency analysis indicated that for many of the scales, and in particular Low Intensity Pleasure, the lowest response category is rarely used. Third, most items from the CBQ EC scales were flagged for revision or omission, many of which demonstrated measurement noninvariance between ages 3-4 and 6-7. This indicates that these items are not related to their parent construct consistently across early childhood. Shortcomings of the CBQ EC scales and items, especially regarding measurement noninvariance, indicate a need for revision. The issues this study uncovers and the argument for revising the CBQ EC scales are discussed in turn below.

Children’s ability to effortfully control behavior, emotions, and attention develops over time (Carlson, 2005; Jones, Rothbart, & Posner, 2003; Rothbart et al., 2003; Rueda, 2012). Therefore, it would be reasonable to suspect that ratings of child effortful control would be higher for older children, and lower for younger children. Interestingly, for the Low Intensity Pleasure scale the opposite appears to be true for several items, and particularly for those on the CBQ-SF. Examining the content of these items, it seems many of these may be more relevant to younger children than older children, as they reflect activities that children may “grow out of” with time. For instance, older children may be more inclined to take showers as opposed to baths (CBQ 54), or be less appreciative of nursery rhymes (CBQ 151) and being sung to (CBQ 146). One item (CBQ 133) shows higher ratings for younger children, possibly because the content of this item refers to looking at “picture books.” Perhaps referring only to “picture books” anchors this item to a period in development when children cannot yet read, and parents of older children may not consider “reading” as equivalent to looking at a picture book. In support of this interpretation, a similar item on the Attentional Focusing subscale (CBQ 186. “Sometimes becomes absorbed in a picture book and looks at it for a long time”) also demonstrated significantly lower ratings for older children ($t$ = -3.63, $p$ < .001, Cohen’s $d$ = -.178).

While Attentional Focusing and Inhibitory Control show increases in ratings across age, Low Intensity Pleasure seems to be moving in the opposite direction. This is concerning for two reasons. First, because we expect effortful control to improve over time, the Low Intensity Pleasure scale and items directly contradict our expectations of the nature of effortful control. Second, because some items are rated lower for older children, this indicates that they may not relate to the construct the same way over time (especially if other items do not change or increase in mean level over time). Comparing these problem items with results from invariance testing, we see that every one of these items demonstrated some form of measurement invariance. Two items (CBQ 54 & 111r) did not relate to the factor ($\lambda$ < .30) for older children, and the remaining items (CBQ 133, 146, 151, and 164) demonstrated scalar non-invariance, meaning that something other than the factor was influencing the item means. What these findings generally indicate is that the Low Intensity Pleasure scale is problematic as a measure (also supported by the response frequency analysis) but also may not relate as closely to the factor of effortful control as previously thought. A similar study showed scalar noninvariance of the low intensity pleasure scale at the factor level by parent (mother vs. father) and child gender, further supporting the need to revise or reconsider this scale as a component of effortful control at this age (Clark et al., 2016). No other EC scales demonstrated this level of noninvariance.

When considering a response use threshold of 5%, it appears that for most items the lower 2-3 response categories are infrequently used. In fact, most EC items have infrequent use in the lowest response category corresponding to “extremely untrue of your child,” and roughly half of all EC items have infrequent use in the second-to-lowest category “quite untrue of your child.” By comparison, most items, but not all, demonstrate sufficient use of the highest response option “extremely true of your child.” Particularly concerning is the Low Intensity Pleasures scale, in which it appears the vast majority of responses use only the highest three options. This may indicate that the items in this scale are generally too “easy” and that more difficult items are needed to better differentiate between children on this construct. Interestingly, nearly all items of the Attention Shifting scale have infrequent use of the lowest and highest response categories. However, this scale does have considerably fewer items on the standard form than the others scales to be used in the response frequency analysis. When examining the response options with negligible use (Table 2.6) it is clear that a considerable portion of most scales have items where the lowest response option is neglected. This is an indication that the rating scale may need to be revised (Penfield, 2013).

Negligible option use may be an artifact of not only the number of options but also the anchors used on the rating scale. DeVellis (2012) recommends using enough response options for respondents to meaningfully discriminate, with the anchoring statements guiding those discriminations. Regarding the CBQ, this raises the question: how well do the anchors discriminate between varying degrees of a temperament trait? Is the difference in value between “extremely untrue” and “quite untrue” the same difference between “quite untrue” and “slightly untrue?” Would a shorter rating scale be able to discriminate as well as a 7-point scale? There is some evidence that 5- and 7-option rating scales perform comparably in terms of reliability (Preston & Coleman, 2000), indicating that an alternative, shorter rating scale may be an adequate alternative for the CBQ. Furthermore, newer measurement approaches such as graded response models (GRM) within an IRT framework may be more easily applied on a scale with fewer response categories, as insufficient cell coverage (i.e., sparse response option use) may bias GRM parameter estimates (Edelen & Reeve, 2007).

Another issue relating to the CBQ rating scale is that the other child temperament questionnaires based on Rothbart's work each use different rating scales. The Early Childhood Behavior Questionnaire (ECBQ; Putnam, Gartstein, & Rothbart, 2006) for toddlers ages 18-36 months uses a 7-point scale with anchors relating to frequency of behavior (e.g., 1 = "never", 2 = "very rarely", etc.) and the Temperament in Middle Childhood Questionnaire (TMCQ; Simonds & Rothbart, 2004) for children ages 7-10 years uses a 5-point scale with anchors mixing both frequency and veracity (1 = "almost always untrue", 2 = "usually untrue", etc.). Because each of these instruments use different rating scales, they may elicit different information from parents—frequency of a behavior may not be the same as a global rating of how "true" a statement relates to their child—and it also requires that items be phrased differently to suit the response options. For example, CBQ 168 states "*My child… can easily stop an activity when s/he is told 'no'*", while ECBQ 111 states *"When told 'no', how often did your child stop an activity quickly?"* and TMCQ 6 states *"My child… can stop him/herself when s/he is told to stop."* Differences in rating scales present challenges for equating and encumber construct comparability.

As EC is a developing aspect of children's temperament, it would be useful to not only be able to measure it consistently over ages 3-7 years (an issue this study attempts to address), but perhaps also examine nascent forms of EC from 18 months of age and earlier through more complex forms at ages 10 and above. Considering temperament is measured differently by the three instruments spanning this age range, our understanding of how temperament and EC change over time is thus hindered by inconsistency with which it is measured. A common rating scale would be the first step toward introducing consistency with ratings, but common items across scales could provide the added benefit for equating the constructs over time in order to adequately study longitudinal change in temperament.

What is apparent from this study is that the CBQ effortful control scales could use revision. A number of items demonstrate measurement noninvariance, many of which have reasonably identifiable sources for differential functioning (e.g., older children are more likely to read rather than look at picture books). Furthermore, several items are outdated (e.g., CBQ 111r. *"Isn't interested in watching quiet TV shows, such as 'Mister Rogers.'"*), are deemed "not applicable" by a large percent of respondents (e.g., CBQ 20 *"Is good at games like 'Simon Says,' 'Mother, May I?' and 'Red Light, Green Light.'"*) or are not salient measures of the construct (e.g., CBQ 12r *"Rarely enjoys just being talked to"* does not load on Low Intensity Pleasure). Some psychometrically sound items that do demonstrate measurement invariance are too close in wording and content (e.g., CBQ 32r *"Has a hard time following instructions"* & CBQ 136 *"Is good at following instructions"*), and thus may be over-represented in the construct they measure.

Revising the CBQ EC scales should provide more consistent measures of effortful control for children ages 3-7, yet the content represented in the EC scales does not fully coincide with how the scales are defined. As mentioned in Chapter 1, both *planning* and *detecting errors* are explicitly incorporated into definitions of effortful control, but these aspects are underrepresented in the CBQ EC items. Furthermore, attention shifting is an underdeveloped component of effortful control, and is absent from the short form instrument. Additional items should be written to address these inconsistencies.

Regarding the rating scale, although reducing the scale from 7-points to 5-points may be an attractive option, I propose the 7-point scale be retained but adopting the anchors from the IBQ-R and ECBQ. The IBQ-R and ECBQ scales require parents rate the *frequency* with which their child displays certain behavior (i.e., Never, Rarely, etc.) and may be preferable to the current CBQ scale by providing a more objective rating of temperament rather than eliciting a global value-judgment. Although the present study shows some infrequent use for response options at the lower-end of the scale (i.e., “Extremely untrue of your child”), this may be remedied at the scale level by introducing more “difficult” items that have more variability among the lower response options. Also, while a revised version of the CBQ should be flexible to alternative measurement approaches (e.g., IRT), workarounds exist for dealing with response sparseness in such frameworks. For example, combining adjacent response options on items with sparse data may be a viable approach (e.g., Edelen & Reeve, 2007), as graded response models can handle item sets that vary in number of response options (Samejima, 2010).

In summary, I propose the CBQ EC scales be revised. I will take a primarily theoretical approach to scale development, although empirical evaluation of items and scales will play a significant role. The revision of these EC scales will involve the aid of content experts, and particularly with those involved in creating other temperament scales, such as Dr. Masha Gartstein and Dr. Samuel Putnam. Using Rothbart’s (Rothbart et al., 2001; Rothbart & Bates, 2006) generally accepted construct definitions and considering scale content, content experts should revise items flagged in this study for demonstrating poor psychometric properties or non-invariance and generate a new item pool. Following the approach taken by Rothbart et al. (2001), parents of young children should be consulted to review, edit, revise, and suggest additional items. Interviews structured to examine item response processes (i.e., cognitive interviews; Knafl et al., 2007) should be used to elicit this feedback, and may help explain why some items operate differently across ages. After finalizing an item pool the items should be piloted with a large sample, and data should be evaluated in a manner similar to that taken in this chapter. Items that demonstrate measurement invariance and have strong psychometric properties (i.e., few “N/A” responses, variability across all or most response options, strong factor loadings), should be selected to comprise the revised scales. However, to ensure the final scales are *theoretically* sound within Rothbart’s framework (e.g., items match the definition, scales measure voluntary aspects of controlled behavior, attention, and emotion, scales are balanced in terms of scale content), they must first be approved by content experts. This will both provide content-related evidence of validity, and some assurance the final scales will not only be based on maximizing the statistical properties using a single sample. As the length of original CBQ scales (12-14 items) has been viewed as prohibitive for some researchers (see: Putnam & Rothbart, 2006) fewer items would be desirable. Putnam and Rothbart (2006) found that eight items was shortest scale length providing adequate psychometric properties for the Low Intensity Pleasure scale, and for consistency I will set the minimum scale length for revised EC scales to eight.

To be considered an improvement on the current versions, the revised CBQ EC scales should achieve the following aims: (a) accurately represent the construct of effortful control as defined by Rothbart (Rothbart et al., 2001; Rothbart & Bates, 2006), (b) show measurement invariance across ages 3-7, and (c) demonstrate strong psychometric properties under multiple measurement frameworks (i.e., CTT, CFA).

# CHAPTER 3

# STUDY 2: REVISING AND PILOTING THE CBQ EFFORTFUL CONTROL SCALES

## Introduction

In this study, I adopt DeVellis' (2017) guidelines for the overall process of scale development, and take two test development approaches for generating items and refining scales: a rational-theoretical approach and a factor-analytic approach. A rational-theoretical approach refers to grounding scale development and item selection in theory, in concert with expert opinion (Hubley & Zumbo, 2013). This approach is similar to how the CBQ was initially constructed (Rothbart et al., 2001), by basing item writing and test development on current theoretical and empirical understanding of the nature of temperament and effortful control (e.g., Rothbart & Bates, 2006; Rueda, 2012). A factor-analytic framework for test development involves using factor analysis or statistical rules to determine what items to retain (Hubley & Zumbo, 2013). Results from Study 1 reflect this approach, as properties of the CBQ EC scales were statistically evaluated and decision rules set to determine which items should be retained and which items should be revised or omitted. Factor analysis will also play a significant role in the present study. The purpose of this study is to revise the CBQ EC scales, pilot an item pool with a large sample of parents, evaluate the items for desired psychometric properties, and finalize a revised set of statistically and theoretically sound effortful control scales (CBQ-R EC).

DeVellis' (2017) provides a series of guidelines for scale development, which I have adapted. This chapter presents a detailed description of each step in the scale development process, which is outlined as here (also see:

Table 3.1). I first recruited several content experts to aid in the scale development process (Step 1). Next, DeVellis recommends clearly defining the construct to be measured and establishing a measurement format (Step 2), so I have worked with content experts to refine the content and definitions of each EC scale and settle on a rating format. DeVellis next suggests generating and refining an item pool (Step 3). Based on the revised definitions, the content experts and I generated an item pool comprised of new items and existing CBQ, ECBQ, and TMCQ EC items and reviewed, edited, and reduced the item pool. At this point I had the items reviewed by another set of experts: parents of children ages 3-7. I conducted individual cognitive interviews with parents to evaluate items and identify remaining issues. Results of these pre-pilot cognitive interviews were discussed with content experts, final edits were made to the item pool, and validation items added. I included two types of validation items: both original CBQ EC scales to provide construct-related evidence of validity, and three items to detect participant inattentiveness to help ensure valid responses. Next, I piloted the items online with a large number of parents of children between ages 3 and 7 (Step 4), and the pilot data was split into a calibration sample and a validation sample. I evaluated the items and refined the scales with the calibration sample response data using both Classical Test Theory and Confirmatory Factor Analysis (CFA) measurement invariance criteria (Step 5). The statistically-refined scales were then reviewed by content experts, and adjustments were made based on theoretical grounds. Finally, I compared the psychometric properties of the original and revised scales using the validation sample response data (Step 6).

Table 3.1

*CBQ-R Effortful Control scale development process, adapted from DeVellis (2017).*

| |
|---|
| Step 1: Recruit Content Experts |
| Step 2: Define Constructs and Measurement Format |
| a. Scale and Content Areas Definitions |
| b. Rating Scale |
| Step 3: Generate and Refine and Item Pool |
| a. Item Writing and Editing |
| b. Pre-Pilot: Cognitive Interviews |
| c. Finalize Pilot Items |
| Step 4: Conduct Pilot |
| Step 5: Evaluate Items and Refine Scales |
| a. Statistical Methods |
| b. Theoretical Refinement |
| Step 6: Compare Original and Revised Scales |

## Step 1: Recruit Content Experts

To ensure content domains are adequately represented, I recruited five temperament experts throughout the revision process: Dr. Samuel Putnam, Dr. Masha Gartstein, Dr. Kathleen Moritz Rudasill, and Dr. Carmen González-Salinas all agreed to participate. Dr. Mary Rothbart also contributed to the selection of items for the final scales. Experts were recruited via email and face-to-face discussion at the Occasional Temperament Conference meeting 2016.

Dr. Putnam (Bowdoin College) has collaborated with Dr. Rothbart on constructing the CBQ-SF scales (Putnam & Rothbart, 2006) and the Early Childhood Behavior Questionnaire (ECBQ; Putnam, Gartstein, & Rothbart, 2006). Dr. Gartstein (Washington State University) revised the Infant Behavior Questionnaire (IBQ-R) with Dr. Rothbart (Gartstein & Rothbart, 2003) and collaborated on producing the ECBQ.

Furthermore, Dr. Putnam and Dr. Gartstein have authored a number of studies examining the measurement and structure of temperament over time using parent report measures (e.g., Putnam, Ellis, & Rothbart, 2001; Putnam, Rothbart, & Gartstein, 2008). Dr. Moritz Rudasill (University of Nebraska-Lincoln) is an expert on the role of temperament in school contexts, and therefore not only provides content expertise but also insight into how questions may be interpreted from a teacher's perspective, as it is common to administer the CBQ to both parents and teachers, though agreement is generally low (Allan et al., 2013; Rudasill et al., 2014). Dr. Carmen González-Salinas (University of Murcia, Spain) has translated the CBQ and most other Rothbart temperament questionnaires into Spanish, and has written extensively on temperament and self-regulation (e.g., González et al., 2001; Carranza, González-Salinas, & Ato, 2013). Dr. González-Salinas is valuable in helping ensure revised EC content represent the regulatory constructs of interest, as well as identify items that may not transcend cultural and geographic boundaries. Finally, Dr. Rothbart (University of Oregon) has been one of the most influential temperament experts over the past several decades, has published extensively on temperament and child development (e.g., Rothbart & Bates, 2006; Rothbart, DerryBerry & Posner, 1994; Rothbart, 1982, 2007), and co-authored the CBQ (Rothbart, et al., 2001), ECBQ-R (Gartstein & Rothbart, 2003), and a number of other instruments for measuring temperament in infancy, childhood, adolescence, and beyond.

## Step 2: Define Constructs and Measurement Format

### Scale and Content Area Definitions

The second step in scale development is creating measurement definitions. For the present study, this involved establishing the aspects of EC to be measured (the scales), their definitions, and fine-grained content areas within each scale. For each existing CBQ EC scale I examined the content of each item and attempted to identify and label content areas that seemed to capture the nature of similar items. For instance, from the Attentional Focusing scale I initially grouped CBQ 171 *"Is easily distracted when listening to a story"* and CBQ 195 *"Has a hard time concentrating on an activity when there are distracting noises"* into the content area "Distractedness." I then distributed these proposed content areas, along with Rothbart's original scale descriptions (Rothbart, Ahadi, Hershey, & Fisher, 2001), to the group of content experts. Each expert individually reviewed the proposed groupings and suggested alternative labels for each item and names for content areas. An additional scale, Activation Control—which appears on the TMCQ—with additional content areas (e.g., "*Persistence*: maintaining attentional focus until task completion") not currently represented in the CBQ were proposed. I compiled the experts' suggestions and as a group via Skype, we decided upon the content areas within each scale, and agreed to include the additional Activation Control scale.

Activation Control was defined as *"the capacity to perform a subdominant action against a natural inclination or aversion,"* and includes the abilities to engage in cognitively, socially, and physically demanding activities (content areas of cognitive activation, social activation, and physical activation) as well as the ability to maintain attentional focus until task completion (persistence). While cognitive, social, and physical activation were based on the items from the TMCQ Activation Control scale (Simonds & Rothbart, 2004), persistence was based on items from the existing CBQ Attentional Focusing scale. Upon reviewing the content of existing items reflecting persistence, the content experts decided the items and content area belonged in Activation Control, as they involve completing an activity that has a clear end-point, which may require staying involved longer than desired (e.g., CBQ 16 *"When picking up toys or doing other jobs, my child usually keeps at the task until it's done."*). Note the definition of Activation Control is slightly different from the original TMCQ definition *"the capacity to perform an action when there is a strong tendency to avoid it."*

The original definition of Attentional Focusing (*"the tendency to maintain attentional focus upon task-related channels"*) was retained from the original CBQ, and the content areas include the intensity of attentional focus (concentration), duration of attentional focus (sustained attention), and the ease with which attention is involuntarily broken (distractibility). These content areas and definitions were based on the original CBQ Attentional Focusing scale.

Attention Shifting was defined as *"the capacity to flexibly shift attention from one activity to another,"* and includes the ability to attend to competing stimuli simultaneously (dividing attention), the ease with which attention is voluntarily broken (release of attention), the ease with which attention is shifted between stimuli or tasks (shifting attention), and the ease with which attentional focus is initiated, including the vigilance of being alert to environmental stimuli (orienting attention).

Inhibitory Control was defined in the original CBQ as *"the capacity to plan and to suppress inappropriate approach responses under instructions or in novel or uncertain situations."* However, upon review of the items comprising the scale, the content experts noted the CBQ item content did not necessarily align with the definition. Although our initial grouping of CBQ inhibitory control items contained the content area "following directions" (e.g., CBQ 32 *"Has a hard time following instructions"*), content experts discussed how following directions could fit into other content areas. For example, one expert noted that following rules when you would prefer not to would be an example of activation control, not inhibitory control. It was also noted that following directions may be confounded with working memory and in some instances, obedience (i.e., following a directive by a parent). Examining original CBQ items that refer to following directions (CBQ 136 *"Is good at following instructions"*, CBQ 168 *"Can easily stop an activity when s/he is told 'no'"*, and CBQ 185 *"Is usually able to resist temptation when told s/he is not supposed to do something"*), each item is absent a context, and therefore it is uncertain how parents are interpreting these statements. The content experts thus decided instead to try to incorporate more specific examples of "following directions" that more clearly align with other subscales during item writing (e.g., a proposed Activation Control item *"If told to do more than one thing (for instance, wash hands then get dressed), my child will do all tasks without being reminded"*). Furthermore, the content experts decided to drop "planning" from the definition of inhibitory control, as it was agreed that planning reflects a secondary aspect of EC and executive functioning. Planning may indeed *require* inhibitory control and EC, but the content experts decided it is not fundamental to the definition or nature of inhibitory control.

The Low Intensity Pleasure scale was controversial among the content experts. Some experts were unsure why LIP was incorporated into EC, arguing that it reflects secondary aspects of EC that—similar to "planning" in IC—merely require the presence of EC. Other experts felt LIP is a core aspect of EC, reflecting a sensitivity and conscientiousness for subtle stimuli, thus facilitating EC. It was recommended I reach out to Dr. Stephan Ahadi, with whom Dr. Mary Rothbart worked to publish the initial CBQ. In our personal correspondence (S. Ahadi, February 6, 2017), Dr. Ahadi indicated that Dr. Rothbart did not initially expect LIP to load with EC, although he suggested "it may be that there are different kinds of low intensity pleasures that are differentially related to effortful control over the course of development." Dr. Ahadi noted that EC is required of certain tasks at young ages. For example, learning to decode words to facilitate reading requires a lot of EC, particularly attention, but as children develop reading skills this process becomes more automated and the role of EC in decoding words diminishes. Dr. Ahadi also recommended differentiating between low intensity pleasures that require attention and those that are more passive. Given this suggestion, we decided to retain LIP as a scale for the time being, and write items which incorporate both high- and low-effort activities. The LIP definition was retained from the original CBQ scale (*"amount of pleasure or enjoyment related to situations involving low stimulus intensity, rate, complexity, novelty, and incongruity"*), and three content areas were established, including enjoyment from quiet activities, enjoyment from listening, and cuddliness.

The original CBQ definition of Perceptual Sensitivity definition (*"amount of detection of slight, low intensity stimuli from the external environment"*) was retained, and the two content areas included a child's propensity to notice changes in the environment, including new objects, small details, and changes in another's physical appearance (visual sensitivity), and detecting subtle details in non-visual sensations, such as quiet sounds and new or unique scents, flavors, or textures (sensory sensitivity). Proposed scale and content areas definitions are presented in Table 3.2.

**Establish Measurement Format**

Because the CBQ EC scale evaluation showed that some response options were rarely used, especially at the low end of the scale, Dr. Gartstein, Dr. Putnam, and I

Table 3.2

*Proposed CBQ-R EC scale and content area definitions.*

**Activation Control:** The capacity to perform a subdominant action against a natural inclination or aversion.

1. ***Cognitive activation:*** the ability to engage in cognitively demanding subdominant actions
2. ***Social activation:*** the ability to engage in socially demanding subdominant actions
3. ***Physical activation:*** the ability to engage in physically demanding subdominant actions
4. ***Persistence:*** maintaining attentional focus until task completion

**Attentional Focusing:** Capacity to maintain attentional focus upon task-related channels.

1. ***Concentration:*** the intensity of attentional focus
2. ***Distractibility:*** the ease with which attention is involuntarily broken
3. ***Sustained Attention:*** the duration of attentional focus

**Attention Shifting:** Capacity to shift attention from one activity to another.

1. ***Dividing attention:*** the ability to attend to competing stimuli simultaneously
2. ***Release of attention:*** the ease with which attention is voluntarily broken
3. ***Shifting attention:*** the ease with which attention is shifted between stimuli or tasks
4. ***Orienting Attention:*** the ease with which attentional focus is initiated, including the vigilance of being alert to environmental stimuli

**Inhibitory Control:** The capacity to suppress or moderate desired behaviors, exert care and caution, and delay actions.

1. ***Inhibiting behavior:*** the ability to suppress a desired behavior, including inhibiting response, slowing motor actions, and lowering voice
2. ***Carefulness:*** the ability to exert care and caution when a situation demands so
3. ***Waiting:*** the ability to delay an action, either in the presence or absence of future gain or reward

**Low Intensity Pleasure:** Amount of pleasure or enjoyment related to situations involving low stimulus intensity, rate, complexity, novelty, and incongruity.

1. ***Enjoyment from quiet activities:*** the amount of enjoyment derived from simple, quiet activities.
2. ***Enjoyment from listening:*** the amount of enjoyment derived from listening to sounds, words, or song
3. ***Cuddliness:*** the amount of enjoyment derived from gentle contact and movement

**Perceptual Sensitivity:** Amount of detection of slight, low intensity stimuli from the external environment.

1. ***Visual sensitivity:*** noticing changes in the environment, including new objects, small details, and changes in another's physical appearance
2. ***Sensory sensitivity:*** detecting subtle details in non-visual sensations, such as quiet sounds, and new or unique scents, flavors, and textures

discussed changing the rating scale. Based on the results of Study 1, I proposed using a 5-point scale rather than a 7-point scale, as this would make adapting CBQ data for item response theory (IRT) applications easier. Dr. Gartstein indicated she preferred the range of responses a 7-point scale provides, and both the ECBQ (Putnam, Gartstein, and Rothbart, 2006) for children ages 1.5 to 3 years, and the IBQ-R (Gartstein & Rothbart, 2003) for infants, use the same 7-point scale: 1=Never, 2=Very Rarely, 3=Less Than Half the Time, 4=About Half the Time, 5=More than Half the Time, 6=Almost Always, and 7=Always. Furthermore, the ECBQ and IBQ-R scales measure the *frequency* of observed behaviors, rather than asking the parent to provide a subjective valuation of how *true* a statement applies to her child as the CBQ does. Frequency ratings may be more desirable because they do not demand parents to judge what their child is *like* but rather what their child actually *does*. Therefore, we opted to retain a 7-point scale but adopt the frequency prompts used in these other temperament measures. Given the interest in studying longitudinal relations of Rothbart's temperament constructs (e.g., Putnam, Ellis, & Rothbart, 2001; Putnam, Rothbart, & Gartstein, 2008) a consistent scale across measurements of temperament would also aid longitudinal construct validity, ensuring that scale scores using different instruments are, at a minimum, both reflections of parents' observations of the frequency of their children's behavior, rather than a mix of frequencies and global valuations.

## Step 3: Generate and Refine an Item Pool

### Item Writing and Editing

Once scale and content area definitions were finalized, I provided the content experts with an item writing document that contained instructions for item writing, the scale name, scale definition, the new rating scale, content areas, and content area definitions. Content experts were asked to write at least two items per content area that (a) fit with the new rating scale (answerable in terms of *frequency* of behavior), (b) provide enough context so parents could think of specific situations, (c) avoid content not applicable to some types of families, cultures, and geographies, and (d) apply to children ages three through seven years of age. I also wrote items for consideration. Content experts wrote a total of 220 items and returned them to me within three weeks.

Once all written items were collected, I combined the new items with 63 original and extra CBQ items, 61 ECBQ items, and 47 TMCQ EC items into a master pool of 391 items. I then edited items for grammatical consistency (for instance, rewording ECBQ items formed as questions to statements). Because original CBQ items are administered with the common stem *"My child:"* which is not included in the item text (e.g., CBQ 122 *"Does not seem to notice parents' facial expressions"*), these items were also reworded to incorporate the stem (e.g., *"My child does not seem to notice parents' facial expressions"*).

Next, I sent out the master pool to content experts to flag items that may not apply to children within our age range, might not make sense or apply in different cultures, or may not fit the rating scale. Forty-eight items were flagged based on age, eight items were flagged based on culture, and 29 items were flagged because they didn't fit the rating scale. Items that were flagged were removed from the item pool, with the exception of CBQ items (45 items for age, six for culture, 26 for rating scale). I then sorted items into scale and content areas and grouped items that seemed to have closely overlapping content (e.g., TMCQ 39 *"Can take a Band-Aid® off when needed, even* *when painful"*, new item *"Will allow parent to remove a splinter if they have one"*, and new item *"Will allow parent to clean a cut or scrape"*).

To avoid content overlap and a large item pool with a number of redundant items, I took several steps to trim the groups of overlapping items. First, I tried to retain all CBQ items that were part of an overlapping item group and were not flagged in my original CBQ Evaluation (see: Chapter 2, Table 2.7). Twenty-five original CBQ items were retained, as well as four extra CBQ items not included in the original long-form scale. Next, I tried to retain unflagged TMCQ and ECBQ items if possible. Eleven original or grammatically modified ECBQ items were retained, as were 15 TMCQ items.

I then either selected one item of each remaining unrepresented item group, or merged multiple items within the group based on content and wording. For example, the items *"My child has trouble finishing the task s/he initiates"*, *"My child quits activities and games before they are complete"*, and "*My child often doesn't complete projects, such as crafts or puzzles, that they had started"* were merged into *"My child begins but doesn't finish activities, such as crafts, puzzles, or games"*. I was cognizant about several features in the items during this selection and merging process. First, I favored items that provided a specific context, rather than those which could be interpreted a number of different ways. For example, I opted to exclude TMCQ 80 *"my child has a hard time paying attention,"* while retaining the new item *"My child has difficulty focusing on an activity or task, such as when listening to a story, learning a new game, or working on a puzzle"*. I also chose to avoid items that seemed too complex. The new item *"My child does not shy away from problems that require mental effort to solve and attempts different strategies (e.g., tries to figure out how pieces fit together, how something* *works),"* was not retained, for example. Items that may be too difficult to understand were also excluded, such as the new item *"My child easily orients to objects and people, engaging attention without hesitation."*

In all, 26 original CBQ items were dropped, as were eight extra CBQ items, 49 ECBQ items, 30 TMCQ items, and 128 newly written items. The pre-pilot list of 150 items was then sent back to content experts to for spell check, content check, and final approval. Six additional items were dropped based on content expert feedback. For example, the new item *"When I'm in another room and my child wants to get my attention, s/he will call out or yell for me instead of coming into the room to talk to me"* was dropped because it was noted this behavior may be confounded by SES in some cultures. The remaining item pool was then used for "pre-pilot" cognitive interviews.

**Cognitive Interviews**

Cognitive interview procedure is an adaptation of the procedure used by Knafl et al (2007), including both think aloud and verbal probing methods. After collecting interviews, limited demographic information and parent interpretations of items were summarized and presented to content experts. We reviewed the item summaries for potentially problematic items, and made decisions about whether items should be retained or dropped.

**Participants.** A convenience, community sample of five parents of young children was recruited for pre-pilot interviews via word of mouth and from their previous participation in a longitudinal study of child development. Parent 1 is a white female with one daughter (age 7 years) that lives at home with her and her husband in a mid-sized college town in Nebraska. Parent 2 is a white male with three daughters, two of whom are ages 4 and 6.5 (the other is 2), that live at home with him and his wife in a mid-sized college town in Nebraska. Parent 3 is a white female with two daughters (ages 4.5 and 7.5 years) that live at home with her and her husband in a mid-sized college town in Nebraska. Parent 4 is a Native American female with two sons (ages 3 and 6.5 years) that live at home with her and her husband in a small city in Nebraska. Parent 5 is an Eastern-European, non-native-English-speaking female with two sons (ages 3.5 and 7) that live at home with her and her husband, also a non-native-English-speaker, in a mid-sized college town in Nebraska.

**Interview Procedure.** The order of the 144 pre-pilot items was randomized, and two versions created with one the reverse order of the other. Interviews lasted between 1 and 1.5 hours, and afterward parents were compensated with a $10 gift card. Parents were informed about the interview purpose and procedures, and then given a packet, including sheet of instructions and the questions. The instructions explained the purpose of the questionnaire itself, the rating scale, and procedures for the cognitive interview (described as a "think aloud"). Parents were then given a sample item to practice the "think aloud" procedure.

We then went through items one-by-one, with the parent reading the item to themselves, and talking about the item aloud. I offered probes throughout the interview, asking about the specific instances they could recall when considering an item, suggestions for alternative wording, if the item applied to them, and so on. I took notes throughout the interview, and in some cases told participants how other parents responded to certain items, to see if they felt the same way. For instance, for the new Item 49 *"My child picks up his/her toys after playing,"* Parent 1 asked if this meant *with* or *without* prompting. She stated her daughter would not pick up toys without prompting, but may with prompting. From then on, after their think aloud with this item, I asked Parents 2-5 if when thinking about the item, they thought about a situation where they did or did not prompt their child to pick up their toys. I then asked them if they would respond differently based on either situation. I transcribed all interviews afterward.

**Results.** I organized the pre-pilot items into four categories based on the comments: major rewrites, minor rewrites, proposed omits, and no changes. Ten items were flagged for major rewrites (Table 3.3), with parent comments indicating the need for content or contextual changes that would likely alter the meaning of the item. Forty-one items were flagged for minor rewrites (Table 3.4), with parent comments indicating the need for grammatical changes that may or may not alter the meaning of the item. For both major and minor rewrite items, I attempted to revise these items myself based on parent feedback and suggestions. Eight items were flagged for omission (Table 3.5) because parents did not agree on an interpretation, stated the item did not apply to their child, or indicated the wording was weird or confusing. For these proposed omits, I was unable to rewrite them to rectify these issues. Furthermore, two additional items flagged for omission were original CBQ items (Item 185 / CBQ 162 *"My child is not very careful and cautious in crossing streets"* and Item 285 / CBQ 151 *"My child likes the sound of words, as in nursery rhymes or poems"*), but these were ultimately retained in the pilot to compare the original CBQ scales against the revised scales. Eighty-three of the original 144 pre-pilot items were not flagged for changes or omission.

Table 3.3

*Pilot item refinement: Items flagged for major revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| **Activation Control** | | | | |
| 16 | Cognitive Activation | | My child takes care of needed tasks, such as brushing teeth, before playing. | My child takes care of needed tasks (e.g., washing hands, going to the bathroom, cleaning up) before playing. |
| 60 | Physical Activation | TMCQ | My child can make him/herself take medicine or eat food that s/he knows tastes bad. | My child takes readily takes medicine even if s/he thinks it will taste bad. |
| 64 | Social Activation | | My child will play with another child, even if that child is not their first choice as a playmate. | When asked to, my child will play with another child, even if my child does not really like them. |
| 8 | Social Activation | | When playing with others, my child stays involved until the end of the game even when losing. | When playing with others, my child stays involved until the end of the game even if s/he is not having much fun. |
| **Attention Shifting** | | | | |
| 91 | Orienting Attention | | My child doesn't respond when someone is trying to get his or her attention. | When my child is focused on something, it is difficult to get their attention. |
| 93 | Orienting Attention | | It takes a long time to get my child engaged in a new activity | It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework. |
| 148 | Distractibility | | My child has a difficult time focusing on play activities when something is going on in the background (e.g., people talking, TV playing). | My child has a difficult time focusing on play activities at home when people are talking in the background. |

Table 3.3, continued

*Pilot item refinement: Items flagged for major revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| **Inhibitory Control** | | | | |
| 11 | Carefulness | | My child tries to be careful when asked to help in the kitchen (e.g., pouring content and stirring). | My child tries to be careful when asked to help with things (e.g., in the kitchen, cleaning) |
| 227 | Inhibiting Behavior | CBQ | My child can easily stop an activity when s/he is told "no." | My child can easily stop a play activity when told it's time to stop. |
| 239 | Inhibiting Behavior | CBQ | My child is able to resist laughing or smiling when it isn't appropriate. | My child will resist laughing when they know it's inappropriate (e.g., at church or similar activity). |

Table 3.4

*Pilot item refinement: Items flagged for minor revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| **Activation Control** | | | | |
| 5 | Following Directions | | My child will follow rules for games, rather than making up their own rules to suit them. | My child follows rules for games, rather than making up their own rules to suit them. |
| 21 | Cognitive Activation | | My child endeavors to learn new skills such as drawing or writing. | My child likes to learn new skills, such as drawing or writing |
| 35 | Persistence | | My child begins but doesn’t finish activities, such as crafts, puzzles, or games. | My child will begin but not finish activities, such as crafts or puzzles. |
| 49 | Physical Activation | | My child picks up her/his toys after playing. | When asked, my child readily cleans up his/her toys after playing. |
| 55 | Physical Activation | TMCQ | My child can take a Band-Aid® off when needed, even when painful. | My child resists taking off Band-Aids® because they're afraid of the pain. |
| 63 | Social Activation | | My child shares his/her favorite toys. | My child shares his/her favorite toys with other children. |
| 68 | Social Activation | TMCQ | My child has a hard time speaking when scared to answer a question. | My child has a hard time speaking when they think they're in trouble. |
| **Attention Shifting** | | | | |
| 73 | Dividing Attention | ECBQ | When playing with a favorite toy, my child will continue to play while at the same time responding to my remarks or questions. | My child responds to my questions even while playing with their favorite toy. |

Table 3.4, continued

*Pilot item refinement: Items flagged for minor revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| 81 | Dividing Attention | | My child is only able to attend to one activity or source of information at a time effectively. | My child can only pay attention to one activity or source of information at a time |
| 82 | Orienting Attention | | My child will pay attention to me right away when I call to him/her. | When playing with a toy, my child will pay attention to me right away when I call to him/her. |
| 96 | Release of Attention | | It is difficult for me to get my child’s attention when s/he is reading or watching TV. | It is difficult for me to get my child’s attention when s/he is focused on reading or watching TV. |
| 104 | Release of Attention | | My child can follow conversations even when playing his/her favorite game. | My child can follow conversations even when playing with his/her favorite toys. |
| 105 | Release of Attention | ECBQ | While playing outdoors, how often did your child look immediately when you pointed at something? | My child will immediately look when I point at something. |
| 107 | Release of Attention | | My child has a difficult time tearing him/herself away from an activity when attention is required elsewhere. | My child has a difficult time tearing him/herself away from a fun activity when attention is required elsewhere. |
| 111 | Shifting Attention | ECBQ | If spoken to when watching TV, my child will reply appropriately then go back to attending to the story. | If spoken to when watching TV, my child will reply appropriately then go back to watching TV. |
| **Attention Focusing** | | | | |
| 125 | Concentration | | When trying to learn how something works (ie., TV remote control, new toy), my child concentrates intensely. | When trying to learn how a new toy works, my child concentrates intensely. |

Table 3.4, continued

*Pilot item refinement: Items flagged for minor revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| 131 | Concentration | | My child has difficulty focusing on an activity or task, such as when listening to a story, learning a new game, or working on a puzzle. | My child has difficulty focusing on an activity or task, such as when listening to a story or working on a puzzle. |
| 142 | Concentration | TMCQ | When participating in an activity, my child has a hard time keeping her/his mind on it. | When participating in crafts or other projects, my child has a hard time keeping her/his mind on it. |
| 156 | Distractibility | CBQ Extra | My child gets distracted from her/his projects when I enter the room. | My child gets distracted from his/her projects (e.g., coloring, art) when I enter the room. |
| 165 | Sustained Attention | CBQ | My child can spend a long time engaged in play with toys, drawing, or coloring. | My child can spend a long time engaged in drawing, coloring, or crafts. |
| 173 | Sustained Attention | | My child entertains him/herself for long periods of time. | While playing alone, my child entertains him/herself for long periods of time. |
| **Inhibitory Control** | | | | |
| 181 | Carefulness | | My child is careful not to spill liquids or foods, like when drinking out of a very full cup, pouring a glass of milk, or carrying glass of water or bowl of soup. | My child is careful to not spill liquids, like when drinking out of a cup, pouring a glass of milk, or carrying a bowl of soup. |
| 190 | Carefulness | | When coloring, my child tries not to paint beyond the line. | When coloring, my child tries not to color beyond the lines. |
| 191 | Carefulness | | My child plays so carelessly that s/he breaks her/his toys. | My child plays rough with their toys and damages them. |

Table 3.4, continued

*Pilot item refinement: Items flagged for minor revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| 195 | Carefulness | | When putting away toys, my child makes sure that they are returned to the right place. | When putting away toys or books, my child makes sure that they are returned to the right place. |
| 205 | Waiting / DoG | CBQ | My child has difficulty waiting in line for something. | My child has a difficult time waiting in line for something fun. |
| 210 | Waiting / DoG | CBQ | My child can wait before entering into new activities if s/he is asked to. | My child can wait before starting a fun activity if s/he is asked to. |
| 220 | Inhibiting Behavior | | My child can keep his/her voice quiet when the situation demands it, such as in a library, during a religious service, or when telling a secret. | My child can keep his/her voice quiet when the situation demands it, such as in a library or during a religious service. |
| 222 | Inhibiting Behavior | | My child has trouble sitting still when s/he is told to (at movies, church, etc.). | My child has trouble sitting still when s/he is told to (e.g., during a religious service). |
| 234 | Inhibiting Behavior | | My child will refrain from taking something s/he wants, like a toy, from a peer | My child will refrain from taking something s/he wants, like a toy, from another child. |
| **Low Intensity Pleasure** | | | | |
| 253 | Cuddliness | | My child does not like getting hugs and kisses from family members. | My child seeks opportunities to get hugs and kisses from family members. |
| 255 | Cuddliness | CBQ | My child enjoys sitting on a caregiver's lap. | My child enjoys sitting on a parent's lap. |
| 266 | Cuddliness | | My child likes getting tucked into his/her bed at night. | My child likes getting tucked into bed at night. |

Table 3.4, continued

*Pilot item refinement: Items flagged for minor revisions based on cognitive interviews.*

| Item | Content Area | Source | Pre-Pilot Item | Proposed Rewrite |
|---|---|---|---|---|
| 270 | Listening | CBQ | My child likes being sung to | My child likes it when a parent or caregiver sings to him/her. |
| 316 | Quiet Activities | | My child enjoys going for walks. | My child enjoys walks outdoors. |
| **Perceptual Sensitivity** | | | | |
| 322 | Scent | | My child notices when a parent is wearing cologne, or has changed his/her fragrance. | My child notices when a parent or other adult is wearing cologne or a fragrance. |
| 330 | Scent | | My child comments when he or she encounters pleasant smells such as perfume, food cooking or candles. | My child comments when he or she experiences nice smells such as perfume, food cooking, or candles. |
| 333 | Sound | ECBQ | My child seems to notice when there are distant sounds, such as from animals or sirens from fire trucks in the distance. | My child seems to notice when there are distant sounds, such as birds or sirens from fire trucks in the distance. |
| 347 | Taste | | My child notices when a meal tastes different than usual (e.g., because of a new recipe, different ingredients, etc.) | My child notices when their food tastes different than usual. |
| 362 | Visual | | My child doesn’t notice little changes in the home decoration. | My child doesn’t notice little changes in the home decoration, or when objects are moved to different places in the house. |
| 374 | Visual | CBQ | My child notices it when caregivers are wearing new clothing. | My child notices it when parents are wearing new clothing. |

Table 3.5

*Pilot item refinement: Items flagged for omission based on cognitive interviews.*

| Item | Scale | Content Area | Source | Pre-Pilot Items |
|---|---|---|---|---|
| 72 | Attention Shifting | Dividing Attention | | My child will pay attention to sounds, such as music, at the same time as looking at something different, such as a toy or book. |
| 198 | Inhibitory Control | Carefulness | TMCQ | My child can stop him/herself from doing things too quickly. |
| 309 | Low Intensity Pleasure | Quiet Activities | | My child has a hard time settling down to watch educational television programs. |
| 319 | Low Intensity Pleasure | Quiet Activities | TMCQ | My child likes to look at trees. |
| 337 | Perceptual Sensitivity | Sound | | My child quickly knows who has come home only by listening their noises. |
| 343 | Perceptual Sensitivity | Sound | TMCQ | My child notices the sound of birds. |
| 350 | Perceptual Sensitivity | Taste | | My child notices flavor differences between two different brands of the same food |
| 388 | Perceptual Sensitivity | Visual | ECBQ | While playing or walking outdoors, my child notices flying or crawling insects? |

I presented the findings to the content experts and on a conference call, Dr Putnam, Dr. Gartstein, Dr. González-Salinas, and I reviewed the major and minor rewrites, and settled on final wording. We made further changes to five major rewrites and six minor rewrites, and decided to keep one omitted item to see how it performed (ECBQ 139 *"While playing or walking outdoors, my child notices flying or crawling insects"*). I met with Dr. Moritz Rudasill four days later to review the interview findings and our decisions on rewrites, and we decided to include one additional rewritten item (*"My child has a hard time focusing on playing when people are talking nearby"*). The rewritten items were then merged into an item pool with the unflagged items.

**Finalizing Pilot Items**

In addition to the refined item pool, I included two types of validation items for the pilot. First, to help establish construct-related validity evidence (DeVellis, 2017), all original CBQ EC scale items were added to the pilot. While 31 original CBQ EC items made it through initial item pool review and cognitive interviews (including five extra CBQ items not typically included in measurement), the remaining 21 original CBQ items were reintroduced to the item pool. One Attentional Focusing item (CBQ 144 *"When building or putting something together, becomes very involved in what s/he s doing, and works for long periods."*) was inadvertently omitted from the pilot, but the rewritten version (Item 165 *"My child can spend a long time engaged in play with toys, drawing, or coloring."*) was included in its place on the scale for evaluation. The original CBQ items would also be used to compare psychometric performance to the new scales to determine if the revision has improved upon the original.

The other validation items were included to determine participant attentiveness. These items asked the participant to select a specific response (e.g., *"For this question, please select 'Very Rarely.'"*). Three such items were incorporated into the pilot item pool, and strategically spread throughout the administered item set. In total, the final pilot item pool contained 30 Activation Control items, 29 Attention Shifting items, 23 Attentional Focusing items, 36 Inhibitory Control items, 25 Low Intensity Pleasure Items, 27 Perceptual Sensitivity items, and three validity check items.

**Step 4: Pilot**

To determine the pilot sample size required for invariance testing, I conducted a power analysis following the approach outlined by MacCallum, Browne, and Sugawara (1996) using SAS software, Version 9.3 of the SAS System for Windows, and syntax provided by Dr. Michael Friendly at York University in Toronto, Canada (http://www.datavis.ca/sasmac/csmpower.html). This analysis estimates the sample size required to achieve adequate power ($\beta$ = .80) within a CFA framework based on the root mean square error of approximation (RMSEA) global fit index. This procedure essentially tests whether a "close" fitting RMSEA value (.05) can be significantly distinguished from a "poor" fitting RMSEA value (e.g., .08, .10, or .12) based on various sample sizes and model degrees of freedom (MacCallum, et al., 1996). Results are presented in Table 3.6, with different CFA models and stringency of RMSEA comparisons organized in rows, and columns organized by number of items (indicators) per analysis (i.e., scale length). Note that sample sizes were rounded up to increments of five.

Table 3.6

*Power analysis to detect differences (β >.80) in global fit for CFA models, by number of scale items.*

| | 12 items | 11 items | 10 items | 9 items | 8 items | 7 items | 6 items |
|---|---|---|---|---|---|---|---|
| Single CFA Model* | df = 54 | df = 44 | df = 35 | df = 27 | df = 20 | df = 14 | df = 9 |
| RMSEA .05 v .12 | 55 | 65 | 75 | 90 | 120 | 140 | 200 |
| RMSEA .05 v .10 | 95 | 120 | 140 | 160 | 200 | 260 | 360 |
| RMSEA .05 v .08 | 220 | 240 | 280 | 360 | 440 | 600 | >600 |
| Configural Invariance Model** | df = 252 | df = 209 | df = 169 | df = 135 | df = 104 | df = 77 | df = 54 |
| RMSEA .05 v .12 | <45 | <45 | <45 | <45 | <45 | 45 | 55 |
| RMSEA .05 v .10 | <45 | <45 | 50 | 55 | 65 | 75 | 95 |
| RMSEA .05 v .08 | 75 | 85 | 95 | 120 | 140 | 160 | 220 |
| Strong Invariance Model*** | df = 287 | df = 241 | df = 199 | df = 161 | df = 127 | df = 97 | df = 71 |
| RMSEA .05 v .12 | <45 | <45 | <45 | <45 | <45 | <45 | 50 |
| RMSEA .05 v .10 | <45 | <45 | 45 | 50 | 55 | 65 | 80 |
| RMSEA .05 v .08 | 70 | 75 | 85 | 100 | 120 | 140 | 180 |

* Single CFA Model df = (k*(k+1)/2+k-3k). No covaried residuals. Factor Mean=0, Variance=1.

** Configural Invariance Model df = (k*(k+1)/2+k-6k). All loadings, intercepts, and residuals estimated for both groups, no covaried residuals. Factor M=0, V=1.

*** Strong Invariance Model df = (k*(k+1)/2+k-3k-1). One set of loadings, intercepts, and residuals estimated and constrained across groups, no covaried residuals. Factor mean estimated. Factor V=1.

Tests of measurement invariance require initially estimating a CFA factor structure for each group independently, and the top section of Table 3.6 displays sample size requirements for single-group CFA models. As can be seen, the most stringent test of RMSEA values requires considerably large sample sizes to distinguish “close” fit and “adequate” or “fair” fit (RMSEA = .08; MacCallum et al., 1996; Browne & Cudeck, 1993). However, a slightly less stringent test between close fit and “mediocre” fit (RMSEA = .10) requires considerably smaller sample sizes. Using this slightly less stringent criteria, models with “close” fit can still be differentiated from poorly fitting models (RMSEA >.10), and therefore this sample size requirement should provide adequate power for the current project. Considering scale length, a single-group CFA with 8 items (indicators) requires a sample size of N=200, whereas a 10-item scale would require a smaller sample of N=140. Note that the two other types of models presented, configural and scalar invariance—I’ve omitted metric invariance, as this model falls between configural and scalar in terms of degrees of freedom—require even smaller samples, as these analyses estimate models with two groups simultaneously therefore using an input covariance structure with twice as many variables, and many more degrees of freedom. Because I identified eight items as the minimum number to include for each scale, and considering the stringency with which I would need to be confident with my model fit, I sought a minimum sample size of N=200 per group (ages 3-4 and 6-7), or N=400 total to adequately evaluate the pilot items.

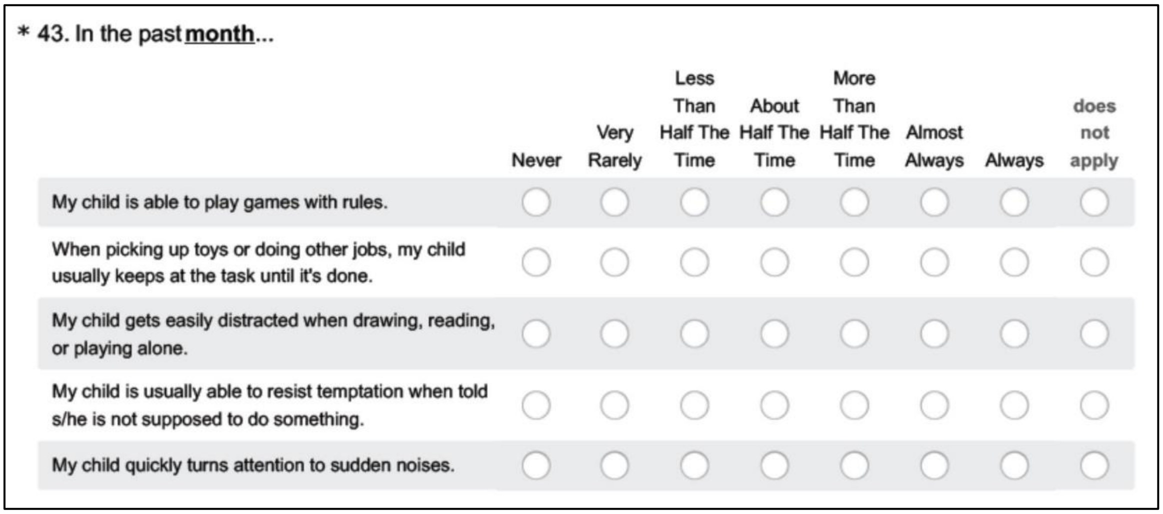
**Questionnaire**

I created the Effortful Control questionnaire online using Survey Monkey (surveymonkey.com). Only parents age 19 and over with typically developing 3 to 7 year-old children who lived at home with them were eligible to participate. Those who completed the questionnaire were entered into a random drawing for one of ten (10) $50 gift cards to Wal-Mart or Target. The questionnaire contained the IRB-approved informed consent form, and asked a few demographic questions of their child (gender, year and month born, ethnicity). Logic was built in to the questionnaire so that parents responded to questions based on only one of their children, in the event they had multiple children age 3-7.

Next, instructions introduced parents to the rating scale, how to respond, and a note about the difference between “Never” and “Does Not Apply.” The instructions also indicated that some questions may sound the same, and that similar questions were included on purpose to determine which wording works best.

The next nine pages contained 170 pilot items plus 3 check items (e.g., “For this question, please select ‘Very Rarely.’”) on pages 3, 5, and 9. Pilot items were randomized, and two versions of the survey were created, where Version 2 contained the pilot items in reverse order of Version 1. Items were grouped into 5-item blocks beginning with the common stem “In the past month…” and scale anchors appearing directly above the corresponding radio buttons of each block (see Figure 3.1 for an example). Participants were required to provide a response to each item.

After the pilot items, parents were asked several demographic questions, as well as if they would be willing to retake the questionnaire in one month. If parents indicated they were willing to retake the questionnaire, they were asked for contact information (phone number or email, and first name). Parents then indicated their preference for gift card in the event they won the drawing (Wal-Mart, Target, Either Wal-Mart or Target) or if they did not want to be entered into the drawing.



* 43. In the past **month**...

| | Never | Very Rarely | Less Than Half The Time | About Half The Time | More Than Half The Time | Almost Always | Always | does not apply |
|---|---|---|---|---|---|---|---|---|
| My child is able to play games with rules. | | | | | | | | |
| When picking up toys or doing other jobs, my child usually keeps at the task until it's done. | | | | | | | | |
| My child gets easily distracted when drawing, reading, or playing alone. | | | | | | | | |
| My child is usually able to resist temptation when told s/he is not supposed to do something. | | | | | | | | |
| My child quickly turns attention to sudden noises. | | | | | | | | |

Figure 3.1
*Sample five-item block from Study 2 online questionnaire (desktop platform).*

**Procedure**

Parents were recruited for participation through Facebook, email, and word of mouth over the course of one month (4/6/2017 through 5/5/2017). I contacted several dozen large parenting-related Facebook pages and groups, and asked for the administrator to post and “pin” a link to the questionnaire with some short introductory text on their page. Three of these pages agreed to post the questionnaire once: Evolutionary Parenting (www.facebook.com/EvolutionaryParenting/; 98,547 followers as of 5/13/17), Mama the Reader (www.facebook.com/MamatheReader/; 1,225 followers as of 5/13/17), and Laura Mazza – Mum on the Run (www.facebook.com/themumontherun/; 16,673 followers as of 5/13/17). One page, Positive Parenting: Toddlers and Beyond (www.facebook.com/PositiveParentingToddlersandBeyond/; 928,578 followers as of 5/13/17) posted a link to the questionnaire on two separate occasions. Of the 672 participants that completed the questionnaire and met inclusion criteria, $N$ = 447 (66.5%)

came from these Facebook groups. Other participants were recruited through personal Facebook shares ($N$ = 94, 14.0%), administrator posts on local child care and elementary school PTO Facebook pages ($N$ = 70, 10.4%), personal posts on consignment, buy-sell-trade, and other Facebook pages ($N$ = 29, 4.3%), administrator posts on the UNL Early Development and Learning Lab Facebook page (www.facebook.com/UNL.EDL; $N$ = 21, 3.1%), and email communications ($N$ = 11, 1.6%).

Participants who indicated a willingness to retake the questionnaire were contacted via email after one month requesting they retake the survey, with a link to a retake survey and a description of the child they previously answered questions about (e.g., “Male child born in December, 2009”). The retake survey included an IRB-approved consent and all original pilot items. Parents also were asked to enter their email and the gender, month and year of birth of their child, in order to match them to their previous data.

On May 31st, 2017, all participants who completed the questionnaire, regardless if they met inclusion criteria for the final sample, were entered into a drawing for one of ten $50 gift cards. Ten participants were randomly selected using a random number generator, contacted via phone or email, and were sent their gift cards.

**Participants**

Of the 1,370 parents who began the questionnaire, a total of 722 answered all pilot items. Only respondents who indicated their primary language was English were included in the final sample for analysis (27 respondents excluded). Respondents who indicated multiple primary languages including English were retained. Parents who reported their child’s birthdate placing them outside the possible range of 36 to 95 months were excluded ($N$ = 23). Finally, the three validation questions were checked, and all remaining parents responded to these correctly. The final sample of respondents ($N$ = 672) was then randomly split into two subsamples using a random number generator for each age group: a calibration sample ($N$ = 400; $N$ = 200 for younger and older age groups each) for refining the scales, and a validation sample ($N$ = 272) for comparing scale properties against the original scales.

## Step 5: Evaluate Items and Refine Scales

### Statistical Refinement

**Participants: Calibration Sample.** The calibration subsample ($N$ = 400) consisted of 200 parents of 3-4 year olds (“younger” age group; 94 female children, mean child age = 46.57 months, SD = 6.83) and 200 parents of 6-7 year olds (“older" age group; 97 female children, mean age = 80.83 months, SD = 7.65). Parents were overwhelmingly female ($N$ = 197 parents of younger children, $N$ = 194 parents of older children). Children in the calibration sample were predominantly white (81.5% and 83.5% for younger and older children, respectively). The remaining children were mixed race (12.5% and 7.5%), Hispanic or Asian (2.0% and 1.5% each), black (0.0% and 2.0%), or some other race (1.5% and 2.5%). Parents of these children were also overwhelmingly white (88.5% and 88.0%), with other parents reporting as mixed race (4.5% in each group), Hispanic (3.0% and 1.5%), Asian (2.0% and 3.5%), Black (0.5% and 1.0%), or some other race (0.5% each). Groups did not differ by child ethnicity, $\chi^2(6) = 7.67$, $p$ = .26, nor parent ethnicity, $\chi^2(6) = 2.49$, p .87. Most parents were married (N=175 and N=173), with the remaining parents either never married ($N$ = 11 and $N$ = 9), separated or divorced ($N$ = 7 and $N$ = 11), some “other” marital status ($N$ = 6 and $N$ = 5), or refused to answer (*N* = 1 and *N* = 3). Groups did not differ by parent marital status, $\chi^2(4) = 1.52$, $p = .82$. The older group reported having significantly more children in the household (M = 2.43, SD = 1.105) compared to the younger group (M = 1.995, SD = .980), $t(399) = 4.21$, $p < .001$, and parents in the older group were significantly older (M = 35.52 years, SD= 5.28) than parents in the younger group (M = 34.49 years, SD = 4.58), $t(397) = 2.08$, $p = .039$. Although the older group had a higher median annual household income (between $90,000 and $99,999) than the younger group (between $80,000 and $89,999), the mean difference was not significant, $t(273) = .04$, $p = .97$. Groups did not differ regarding the questionnaire version, $\chi^2(1) = 1.58$, $p = .21$, although more parents took Version 2 (*N* = 123 and *N* = 136) than took Version 1 (*N* = 77 and *N* = 64).

Nearly all parents in the older group reside in the United States (*N* = 183) with some living in Canada (*N* = 8) or elsewhere (*N* = 9). The two age groups differed on country of residence, $\chi^2(2) = 24.65$, $p < .001$, with slightly more parents of younger children residing in the UK (*N* = 20) than parents of the older group (*N* = 3). Most parents in the younger group also reside in the United States (*N* = 148), but some reside in Canada (*N* = 11) and a sizeable number live elsewhere (N = 41; nearly half of whom live in the United Kingdom, N = 20). Using participant-provided ZIP Codes, participants in the U.S. were categorized as residing in one of three types of urban or rural areas: Urbanized Areas, in which at least 50,000 people reside in a geographical area; Urban Clusters, geographical areas containing at least 2,500 but fewer than 50,000 people, and Rural Areas, where fewer than 2,500 people reside (www.census.gov). Geographical location information is based on the 2010 US Census data, and accessed from MABLE/Geocorr12 Geographic Correspondence Engine at the Missouri Census Data

Center website (http://mcdc.missouri.edu/websas/geocorr12.html). The younger and older groups did not significantly differ in geographical area of residence, $\chi^2(2) = 3.49$, $p = .17$, with households located in primarily Urbanized Areas ($N = 122$ and $N = 133$), with some located in Urban Clusters ($N = 15$ and $N = 23$) and Rural Areas ($N = 15$ and $N = 23$).

**Analysis Procedure.** The calibration sample was used to refine the scales. I employed several decision rules for retaining items with the intent to reduce scales to a minimum of eight items, ensuring all content areas were represented and the scales display strong measurement invariance. Items with more than 10% "does not apply" responses will be dropped (Decision Rule 1). The CBQ EC evaluation in Chapter 2 flagged items with 20% or more NA responses for either group, coinciding with decision rules during the construction of the CBQ-SF (Putnam & Rothbart, 2006). However, with a large item pool I was more selective in inclusion criteria, and chose a stricter threshold with which to retain items. Items with 3 or more response categories with negligible response data (< 1%) were dropped (Decision Rule 2). This is the same as the Chapter 2 evaluation decision rule, based on a recommendation by Penfield (2013). Remaining items were included in an CTT item analysis, and items with corrected item-total correlation values less than .30 were omitted (Decision Rule 3) in order to reduce the item pool further before factor analysis evaluation.

Remaining items were included in a multi-group configural CFA, with loadings and intercepts freely and simultaneously estimated for each younger and older group. To reduce redundancy and avoid capitalizing on inter-item correlations due to similar wording or context, modification indices were consulted to identify items pairs with residuals that would correlate above and beyond that explained by the factor. The minimum modification index value was set to 6.635, which indicated undertaking the modification would correspond to a change in the model $\chi^2$ statistic significant at $p < .01$. The item pairs with the largest modification index were evaluated on several aspects to decide which item should be dropped and which retained (Decision Rule 4). These aspects were: (a) factor loadings, with higher loadings preferable, (b) difference in factor loadings between young and old group, with smaller differences preferable, as items with larger differences may fail tests of metric noninvariance, (c) item means, with means closer to 4 (the center of the rating scale) preferable (DeVellis, 2017), (d) item variances, with larger variances preferable (DeVellis, 2017), (e) representation of content areas, with preference to retain a balance of content representation, or at least one item per content area, and (f) the number of other items modification indices recommend for covarying residuals, with fewer preferable. Regarding aspect (f), for example, assume Item X and Item Y show a large modification index for a covaried residual—they are highly correlated beyond what's explained by the factor—if the modification indices also recommend covarying residuals between Item X and five other items (Item A-Item E) but only covarying residuals between Item Y and one other item (Item F), the preference would be to drop Item X. This is because dropping Item X resolves more instances of items overlapping above and beyond the relationship to the factor. Configural models were re-estimated after dropping each item, until no modifications at or above values of 6.635 were recommended or the model fit adequately according to RMSEA, $\chi^2$, and CFI.

Next, I examined the standardized loadings for both groups in the configural model, and dropped items if loadings for both groups were less than or equal to .400 (Decision Rule 5). Although the corresponding decision rule from Chapter 2 set this threshold at .300, I was more selective with a large item pool, and chose to set this value higher. The multi-group model was then checked to verify loadings for each group were significant, and I dropped items with non-significant loadings (Decision Rule 6).

The final configural model was the base model used for invariance testing, the same procedure used in Chapter 2. Metric invariance (Decision Rule 7) and scalar invariance (Decision Rule 8) were tested and items which demonstrated measurement noninvariance were dropped. If only 8 items remained in the scale when an item showed noninvariance, the loading or intercept was freed to improve model fit. Nested models were evaluated using the same three criteria as in Chapter 2: non-significant ($p < .05$) loglikelihood difference test (ΔLL; Brown, 2006), change in Comparative Fit Index (ΔCFI) less than .002 (Meade et al., 2008), and change in Root Mean Square Error of Approximation (ΔRMSEA) less than .015 (Meade et al., 2008). Models passing two of these three criteria were considered comparable to the model of the previous step (e.g., metric model compared to configural model). A list of all decision rules can be found in Table 3.7.

For comparison purposes, I also evaluated the original CBQ scales for measurement invariance using the calibration sample. I initially estimated a multi-group model, and dropped any items that demonstrated configural noninvariance. I did not covary residuals between items. I then further tested for noninvariance in the same manner as described above, and I freed loadings or intercepts for items which demonstrated metric or scalar noninvariance, respectively.

Once all scales were set, I presented the revised scales to the content experts to get final approval on the items. Content experts recommended changes to the scales

Table 3.7

*Decision rules for statistically evaluating items and refining CBQ Effortful Control scales.*

**Descriptives**

1. Missing Data
   - Item dropped from subsequent analyses if either group had ≥10% responses as “Does Not Apply”
2. Negligible Response Data
   - Item dropped if either group had ≥ 3 response options with negligible use (<1% of sample selected the category)
3. Item-Total Correlation
   - Item dropped if correlation between item and corrected total (without item) was < .30

**Factor Analysis (multi-group CFA)**

4. Covaried Residuals
   - Item pairs with modification indices proposing covarying residuals were compared on:
     i. Factor loadings (higher preferable),
     ii. Difference in factor loadings between Young and Old group (smaller preferable),
     iii. Item means (closer to 4 preferable),
     iv. Item standard deviationss (larger preferable)
     v. Number of other items suggested for covarying residuals (fewer preferable)
     vi. Representation of content area (prefer to retain a balance of content area representation, or at least 1 item per content area)
5. Low Loadings
   - Items with loadings for both groups < .40 were dropped, if possible.
6. Configural Noninvariance
   - Items with non-significant factor loadings for either group were dropped, if possible.
7. Metric Noninvariance
   - Items with factor loadings that could not be constrained across groups were dropped, if possible.
8. Scalar Noninvariance
   - Items with intercepts that could not be constrained across groups were dropped, if possible.

based on theory, and I made adjustments to scales until a consensus was reached. From there I evaluated the theoretically-revised scales by testing for measurement invariance using the calibration sample, and examined scale internal consistency and factor structure using the validation sample.

**Scale Refinement.** To evaluate items based on Decision Rules 1-3 (missing data, negligible response data, item-total correlations), data from the calibration sample were analyzed using R version 3.4.0 (R Core Team, 2017). The "alpha" function from R package "psych" version 1.7.3 (Revelle, 2017) was used to examine response frequencies and corrected item-total correlations.

Based on Decision Rule 1, eight items were dropped from consideration for having 10% or more NA responses for either age group, four of which were original CBQ items (Item 238 / CBQ 20 *"My child is good at games like 'Simon Says,' 'Mother, May I?' and 'Red Light, Green Light.'"*; Item 239 / CBQ 116 *"My child is able to resist laughing or smiling when it isn't appropriate."*; Item 308 / CBQ 111 *"My child isn't interested in watching quiet TV shows, such as 'Mister Rogers.'"*; Item 338 / CBQ 105 *"My child usually comments if someone has an unusual voice."*), and the other four new items (Item 64, *"When asked to, my child will play with another child, even if my child does not really like them."*; Item 120 *"My child adapts quickly when the rules of the game change."*; Item R05 *"My child will resist laughing when they know it's inappropriate (e.g., at church or similar setting)."*; Item 322 *"My child notices when a parent or other adult is wearing cologne or a fragrance."*). Based on Decision Rule 2, six additional items were dropped because of negligible use of 3+ response categories in either age group, with one original CBQ item (with a slight rewrite) dropped (Item 293 /

CBQ 133 *"My child enjoys looking at or reading books"*), and five new items dropped (Item 88 *"My child reacts quickly when hearing a song that s/he likes.";* Item 196 *"My child pets and touches animals softly, rather than roughly."*; Item 266 *"My child likes getting tucked into bed at night."*; Item 281 *"My child enjoys having an adult read to them."*; Item 316 *"My child enjoys walks outdoors."*). All of these items had negligible response use at the lower end of the scale (i.e., response categories "Never", "Very Rarely", "Less than Half the Time"). Between rules 1 and 2, one Activation Control item was dropped (29 remained), as were two Attention Shifting items (27 remained), zero Attentional Focusing Items (23 remained), four Inhibitory Control items (32 remained), five Low Intensity Pleasure items (20 remained), and two Perceptual Sensitivity items (25 remained).

The, remaining items for each scale were entered into an item analysis, and items with the lowest corrected item-total correlations below .30 were iteratively removed until all remaining scale items held corrected item-total correlations of .30 or larger. Seven Activation Control items were dropped (22 remained), as were two Attention Shifting items (25 remained), two Attentional Focusing items (21 remained), four Inhibitory Control items (28 remained), six Low Intensity Pleasure items (14 remained), and four Perceptual Sensitivity items (21 remained).

To evaluate items based on Decision Rules 4-8 (covaried residuals, factor loadings, and configural, metric, and scalar noninvariance), multi-group CFA models using data from the calibration sample were estimated using Mplus version 7.11 (Muthén & Muthén, 1998-2013) with robust maximum likelihood estimation (MLR).

***Activation Control.*** A configural multi-group CFA was estimated with the remaining 22 Activation Control items. Modification indices for covarying residuals to improve model fit were consulted, and item pairs with large modification values were evaluated to determine which item to keep. For instance, modification indices indicated a strong need to covary residuals between Item 29 *("My child will keep working on a puzzle or project until it is complete"*) and Item 35 (*"My child begins but does not finish activities, such as crafts or puzzles."*, reversed). Item 35 held larger standardized factor loadings ($\lambda_{young}$ = .652, $p < .001$; $\lambda_{old}$ = .707, $p < .001$) than item 29 ($\lambda_{young}$ = .591, $p <$ .001; $\lambda_{old}$ = .625, $p < .001$), and had mean values nearer to 4.0 (*Myoung* = 4.55, *Mold* = 4.87) compared to item 29 (*Myoung* = 4.79, *Mold* = 5.02), while both items had similar variances ($\sigma^2_{young}$ = 1.51, $\sigma^2_{old}$ = 1.46, versus $\sigma^2_{young}$ = 1.74, $\sigma^2_{old}$ = 1.54). Thus item 35 was retained over item 29 at this step. Twelve Activation Control items were dropped here according to Decision Rule 4.

A configural model was estimated with the remaining 10 items, and based on Decision Rule 5, a single item (Item 61 *"My child has a hard time getting moving when tired."*, reversed) was dropped from the Activation Control scale due to loadings < .40 for both groups. All remaining nine items estimated in a configural model had significant loadings at $p < .001$, so no items were dropped based on Decision Rule 6 (configural noninvariance). This configural model fit well, $\chi^2(54) = 50.67$, $p = .60$, RMSEA = .000 (90% CI: .000 – .040), CFI = 1.000.

Next, all factor loadings were constrained to be equal across groups. This metric model fit significantly worse than the configural model, $\Delta$LL $\chi^2(9) = 25.47$, $p = .0025$, $\Delta$RMSEA = .031, $\Delta$CFI = -.021. Modification indices indicated that freeing the loading of Item 33 (*"My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them."*) would improve model fit the most, so this item was dropped for displaying metric noninvariance. A new configural model was estimated, which displayed good fit, $\chi^2(40) = 35.32$, $p = .68$, RMSEA = .000 (90% CI: .000 – .040), CFI = 1.000. A metric model was subsequently estimated by constraining the loadings of this configural model to be equal across groups. Again, the metric model fit significantly worse than the configural model, ΔLL $\chi^2(8) = 17.88$, $p = .02$, ΔRMSEA = .021, ΔCFI = -.010. Because the minimum number of items was set to eight (based on the prior power analysis), rather than dropping additional items that demonstrated noninvariance, the modification index recommendations were consulted to free item loadings until the metric model fit comparably to the configural model. The loading for Item 43 (*"My child quickly gives up on an activity or project when s/he becomes bored of it."*, reversed) was freed, with item loading on the factor stronger in the older group (unstandardized $\lambda = 1.030$) compared to the younger group (unstandardized $\lambda = 0.649$). This model fit comparably to the configural model, ΔLL $\chi^2(7) = 8.83$, $p = .27$, ΔRMSEA = .000, ΔCFI = .000, and therefore was retained. A scalar model was then estimated by constraining item intercepts (with the exception of Item 43) across groups and freeing the factor mean for the older group. This scalar model fit comparatively worse than the metric model, ΔLL $\chi^2(6) = 35.04$, $p < .0001$, ΔRMSEA = .048, ΔCFI = -.059. Intercepts for Item 5 (*"My child follows rules for games, rather than making up their own rules to suit them."*), Item 9 (*"When I'm busy, my child can find another activity to do when asked."*), and Item 56 (*"My child follows daily routines on his/her own, such as brushing his/her teeth."*) were iteratively freed, and the remaining revised Activation Control scalar model with eight items fit comparatively well to the metric model, ΔLL $\chi^2(3) = 2.81$, $p = .42$, ΔRMSEA = .000, ΔCFI = .000, and overall fit very well, $\chi^2(50) = 46.72$, $p = .61$, RMSEA = .000 (90% CI: .000 - .041), CFI = 1.000.

***Attention Shifting.*** A configural multi-group CFA was estimated with the remaining 25 Attention Shifting items. Modification indices for covarying residuals to improve model fit were consulted, and item pairs with large modification values were evaluated to determine which item to keep. For instance, modification indices indicated a strong need to covary residuals between Item 91 (*"When my child is focused on something, it is difficult to get their attention."*, reversed) and Item 94 (CBQ 6 *"It is hard to get my child's attention when s/he is concentrating on something."*, reversed). Both items had similar standardized factor loadings ($\lambda_{young} = .740$, $p < .001$ & $\lambda_{old} = .743$, $p < .001$ for Item 91, versus $\lambda_{young} = .765$, $p < .001$; $\lambda_{old} = .767$, $p < .001$ for Item 94), similar means ($M_{young} = 4.13$, $M_{old} = 4.18$ for Item 91 versus $M_{young} = 4.18$, $M_{old} = 4.22$ for item 94), similar variances ($\sigma^2_{young} = 1.72$, $\sigma^2_{old} = 2.02$ for Item 91, versus $\sigma^2_{young} = 1.54$, $\sigma^2_{old} = 1.91$ for Item 94). Because Item 94 is an original CBQ item, it was retained over Item 91. Eleven Attention Shifting items were dropped here according to Decision Rule 4, while 14 items remained.

All remaining 14 items estimated in a configural model had loadings of at least .40 in either group, and were all significant at $p < .001$, so no items were dropped based on Decision Rule 5 (low loadings) or Decision Rule 6 (configural noninvariance). This configural model fit well, $\chi^2(154) = 194.77$, $p = .012$, RMSEA = .036 (90% CI: .017 – .051), CFI = .969.

Next, all factor loadings were constrained to be equal across groups. This metric model fit comparably to the configural model, ΔLL $\chi^2(14) = 14.71$, $p = .40$, ΔRMSEA = -.001, ΔCFI = -.001, and therefore was retained. A scalar model was then estimated by constraining item intercepts across groups and freeing the factor mean for the older group. This scalar model fit comparatively worse than the metric model, ΔLL $\chi^2(13) = 108.78$, $p < .0001$, ΔRMSEA = .025, ΔCFI = -.067. Four items demonstrated varying degrees of scalar noninvariance, and each were iteratively dropped, with configural, metric, and scalar models estimated after each. A final scalar model was estimated with all items demonstrating scalar invariance. However, one item present in this scalar model held low factor loadings (Item 118, *"After having been interrupted during an activity, my child has difficulty returning to the activity."*, reversed; standardized $\lambda_{young} = .384$, $\lambda_{old} = .393$), and was dropped. Configural, metric and scalar models were again estimated, with each fitting comparably to the previous model. The final scalar invariance model for the revised Attention Shifting scale contained 9 items and fit very well, $\chi^2(71) = 83.51$, $p = .15$, RMSEA = .030 (90% CI: .000– .053), CFI = .984.

For the original CBQ Attention Shifting scale, two items demonstrated scalar noninvariance (Item 99 / CBQ 180 *"My child has an easy time leaving play to come to dinner."* and Item 115 / CBQ 29 *"My child can easily shift from one activity to another."*). The final scalar invariance model for the original five-item CBQ Attention Shifting scale fit poorly, $\chi^2(17) = 45.44$, $p < .001$, RMSEA = .091 (90% CI: .060 – .124), CFI = .930.

***Attentional Focusing.*** A configural multi-group CFA was estimated with the remaining 21 Attentional Focusing items. Modification indices for covarying residuals to improve model fit were consulted, and item pairs with large modification values were evaluated to determine which item to keep. For instance, modification indices indicated a strong need to covary residuals between Item 143 (CBQ 171*"My child is easily distracted when listening to a story."*, reversed) and Item 180 (*"While listening to a long story, my child will stay focused the whole time."*). Both items had comparable standardized factor loadings ($\lambda_{young}$ = .490, $p$ < .001; $\lambda_{old}$ = .565, $p$ < .001 for Item 143, $\lambda_{young}$ = .532, $p$ < .001; $\lambda_{old}$ = .504, $p$ < .001 for Item 180), although Item 180 had a smaller difference in standardized loading size between groups than Item 143 (.028 versus .075). Both items had similar means ($M_{young}$ = 5.31, $M_{old}$ = 5.32, versus $M_{young}$ = 5.26, $M_{old}$ = 5.39 for Item 180), but Item 180 had more variability ($\sigma^2_{young}$ = 1.72, $\sigma^2_{old}$ = 1.62) compared to Item 143 ($\sigma^2_{young}$ = 1.17, $\sigma^2_{old}$ = 1.35). However, I judged the items to be similar enough to retain the original CBQ item (Item 143). Ten Attentional Focusing items were dropped here according to Decision Rule 4.

A configural model was estimated with the remaining 11 items, and based on Decision Rule 5, a single item (Item 162 *"My child sometimes becomes absorbed in a picture book and looks at it for a long time."*) was dropped from the Attentional Focusing scale due to low loadings for both groups ($\lambda_{young}$ = .377, $p$ < .001; $\lambda_{old}$ = .289, $p$ < .001). All remaining ten items estimated in another configural model had significant loadings at $p$ < .001, so no items were dropped based on Decision Rule 6 (configural noninvariance). A new configural model was estimated, which displayed good fit, $\chi^2(40)$ = 35.32, $p$ = .68, RMSEA = .000 (90% CI: .000 – .040), CFI = 1.000.

Next, all factor loadings were constrained to be equal across groups. This metric model fit comparably to the configural model, ΔLL $\chi^2(10)$ = 10.05, $p$ = .44, ΔRMSEA = -

.003, ΔCFI = -.001. A scalar model was then estimated by constraining item intercepts across groups and freeing the factor mean for the older group. This scalar model fit comparatively worse than the metric model, ΔLL $\chi^2(9) = 51.85$, $p < .0001$, ΔRMSEA = .016, ΔCFI = -.045. Modification indices suggested freeing intercepts for Item 178 (*"My child will stay with the same activity for a half-hour or more if given the chance."*) and Item R01 (*"My child spends a long time engaged in drawing, coloring, or crafts."*), with Item R01 having a slightly larger modification index than Item 178. A set of models was estimated after dropping Item R01, and with the scalar model still fitting worse than the metric model, Item 178 was also dropped and another set of models with the remaining eight items was estimated. Configural, metric, and scalar models were estimated, each fitting comparably to the prior model, and the final scalar model fit well, $\chi^2(55) = 76.53$, $p = .029$, RMSEA = .044 (90% CI: .015– .067), CFI = .966.

For the original CBQ Attentional Focusing scale, one item demonstrated configural noninvariance (Item 41 / CBQ 160 *"My child has difficulty leaving a project s/he has begun."*), with loadings for neither group significant, and three items demonstrated scalar noninvariance (Item 40 / CBQ 47 *"My child will move from one task to another without completing any of them."*, Item 127 / CBQ 125 *"When drawing or coloring in a book, my child shows strong concentration."*, and Item R01 / CBQ 144 rewrite *"My child spends a long time engaged in drawing, coloring, or crafts."*). The final scalar invariance model contained eight of the original nine CBQ Attentional Focusing items—one was dropped due to configural noninvariance—and fit poorly, $\chi^2(52) = 121.59$, $p < .001$, RMSEA = .082 (90% CI: .063 – .101), CFI = .873.

***Inhibitory Control.*** A configural multi-group CFA was estimated with the remaining 27 Inhibitory Control items. Modification indices for covarying residuals to improve model fit were consulted, and item pairs with large modification values were evaluated to determine which item to keep. For instance, modification indices indicated a strong need to covary residuals between Item 6 (CBQ 32 *"My child has a hard time following instructions."*, reversed) and Item 7 (CBQ 136 *"My child is good at following instructions."*). Both items had similar standardized factor loadings ($\lambda_{young}$ = .718, $p$ < .001 and $\lambda_{old}$ = .672, $p$ < .001 for Item 6, and $\lambda_{young}$ = .690, $p$ < .001 and $\lambda_{old}$ = .694, $p$ < .001 for Item 7). However, Item 6 had mean values nearer to 4.0 ($M_{young}$ = 4.86, $M_{old}$ = 5.21) compared to Item 7 ($M_{young}$ = 4.97, $M_{old}$ = 5.32), as well as more variability ($\sigma^2_{young}$ = 1.23, $\sigma^2_{old}$ = 1.25, versus $\sigma^2_{young}$ = 1.02, $\sigma^2_{old}$ = 1.17). Thus Item 6 was retained over Item 7 at this step. Sixteen Inhibitory Control items were dropped here according to Decision Rule 4.

A configural model was estimated with the remaining 12 items, all of which had significant standardized loadings greater than .40 in both groups, so no items were dropped based on Decision Rule 5 or Decision Rule 6. This configural model fit very well, $\chi^2$(108) = 109.36, $p$ = .45, RMSEA = .008 (90% CI: .000 – .037), CFI = .999.

Next, all factor loadings were constrained to be equal across groups. This metric model fit significantly worse than the configural model based on two of the three criteria, ΔLL $\chi^2$(12) = 20.70, $p$ = .055, ΔRMSEA = .012, ΔCFI = -.007. Modification indices indicated that freeing the loading of Item 217 (*"My child has a hard time waiting his/her turn to talk when excited."*) would improve model fit, so this item was dropped for displaying metric noninvariance. A new configural model was estimated, which displayed good fit, $\chi^2(88) = 94.22$, $p = .31$, RMSEA = .019 (90% CI: .000– .044), CFI = 994. A metric model was subsequently estimated by constraining the loadings of this configural model to be equal across groups. This model fit comparably to the configural model, ΔLL $\chi^2(11) = 13.12$, $p = .29$, ΔRMSEA = .001, ΔCFI = .992, and therefore was retained. A scalar model was then estimated by constraining item intercepts across groups and freeing the factor mean for the older group. This scalar model fit comparatively worse than the metric model, ΔLL $\chi^2(10) = 42.40$, $p < .0001$, ΔRMSEA = .022, ΔCFI = -.030. Modification indices indicated scalar noninvariance for Item 200 (*"My child waits his/her turn in games."*), so that item was dropped. Configural, metric, and scalar models for the remaining 10 items fit comparatively well compared to the prior model, and the final scalar model fit very well, $\chi^2(89) = 102.24$, $p = .16$, RMSEA = .027 (90% CI: .000 - .049), CFI = .984.

For the original CBQ Inhibitory Control scale, two items demonstrated scalar noninvariance (Item 238 / CBQ 20 *"My child is good at games like "Simon Says," "Mother, May I?" and "Red Light, Green Light."* and Item X01 / CBQ 63 *"My child prepares for trips and outings by planning things s/he will need."*). The final scalar invariance model for the original thirteen-item CBQ Inhibitory Control scale fit adequately, $\chi^2(153) = 261.21$, $p < .001$, RMSEA = .059 (90% CI: .047– .072), CFI = .918.

***Low Intensity Pleasure.*** A configural multi-group CFA was estimated with the remaining 14 Low Intensity Pleasure items. However, most loadings were below .40, and after initially removing items with covaried residuals and low loadings, fewer than eight items remained. Because of this, all items dropped from Decision Rule 3 (low item-total correlations) were reintroduced into the remaining Low Intensity Pleasure item pool, and I decided to loosen my criteria for Decision Rule 5 to drop items with factor loadings less than .30. Twenty items were then included in another configural multi-group CFA, and I consulted modification indices for covarying residuals to improve model fit. Item pairs with large modification values were evaluated to determine which item to keep. For instance, modification indices indicated a strong need to covary residuals between Item 247 (CBQ 76 *"My child enjoys "snuggling up" next to a parent."*, reversed) and Item 255 (CBQ 174 *"My child enjoys sitting on a parent's lap."*). Both items had similar standardized factor loadings ($\lambda_{young}$ = .449, $p$ < .001 and $\lambda_{old}$ = .718, $p$ < .001 for Item 247, and $\lambda_{young}$ = .505, $p$ < .001 and $\lambda_{old}$ = .721, $p$ < .001 for Item 255). However, Item 255 had mean values nearer to 4.0 ($M_{young}$ = 5.78, $M_{old}$ = 5.60) compared to Item 247 ($M_{young}$ = 6.28, $M_{old}$ = 6.15), as well as more variability ($\sigma^2_{young}$ = 1.45, $\sigma^2_{old}$ = 2.19, versus $\sigma^2_{young}$ = 0.85, $\sigma^2_{old}$ = 1.30). Item 255 also had a smaller difference in factor loadings (.216 versus .269). Thus Item 255 was retained over Item 247 at this step. Six Low Intensity Pleasure items were dropped here according to Decision Rule 4. Based on the loosened Decision Rule 5 criteria, three additional item were iteratively dropped: Item 306 (*"My child enjoys calm inside activities such as building blocks, Legos, dolls, puzzles, etc."*; $\lambda_{young}$ = .281, $\lambda_{old}$ = .168,), Item 277 (*"My child doesn't enjoy being read to very much."*; $\lambda_{young}$ = .201, $\lambda_{old}$ = .200), and Item 282 (*"My child rarely enjoys just being talked to."*; $\lambda_{young}$ = .140, $\lambda_{old}$ = .259). A configural model was estimated with the remaining 11 items, which fit well, $\chi^2(88)$ = 112.90, $p$ = .04, RMSEA = .038 (90% CI: .010 – .057), CFI = .936.

Next, all factor loadings were constrained to be equal across groups. This metric model fit comparatively well to the configural model, ΔLL $\chi^2(11)$ = 10.53, $p$ = .48,

$\Delta$RMSEA = -.003, $\Delta$CFI = .002. A scalar model was then estimated by constraining item intercepts across groups and freeing the factor mean for the older group. This scalar model fit comparatively worse than the metric model, $\Delta$LL $\chi^2(10) = 76.03$, $p < .0001$, $\Delta$RMSEA = .025, $\Delta$CFI = -.142. Modification indices indicated scalar noninvariance for Item 298 (*"My child enjoys drawing and coloring."*), and Item 251 (*"My child enjoys it when I softly rub his/her back."*). First, Item 298 was dropped and new sets of models were estimated, but the subsequent scalar model was still a comparatively poorer fit to the metric model, so Item 251 was dropped and a new set of models estimated. The metric model fit as well as the configural model, and the scalar model fit as well as the metric model. The final scalar model with 9 items fit the data well, $\chi^2(71) = 95.356$, $p$ = .03, RMSEA = .041 (90% CI: .014 - .062), CFI = .913.

For the original CBQ Low Intensity Pleasure scale, five items demonstrated configural noninvariance (Item 277 / CBQ 66 *"My child doesn't enjoy being read to very much."*, Item 282 / CBQ 12 *"My child rarely enjoys just being talked to."*, Item 285 / CBQ 151 *"My child likes the sound of words, as in nursery rhymes or poems."*, Item 293 / CBQ 133 *"My child enjoys looking at or reading books."*, Item 303 / CBQ 86 *"My child doesn't care much for quiet games."*). The final scalar invariance model contained eight of the original 13 CBQ Low Intensity Pleasure scale items—the configurally noninvariant items were dropped—and fit poorly, $\chi^2(55) = 101.48$, $p < .001$, RMSEA = .065 (90% CI: .045 – .085), CFI = .881.

***Perceptual Sensitivity.*** A configural multi-group CFA was estimated with the remaining 21 Perceptual Sensitivity items. Modification indices for covarying residuals to improve model fit were consulted, and item pairs with large modification values were evaluated to determine which item to keep. For instance, modification indices indicated a strong need to covary residuals between Item 377 (CBQ 28 *"My child usually doesn't comment on changes in parents' appearance.",* reversed) and Item 380 (CBQ 65 *"My child comments when a parent has changed his/her appearance."*). Item 380 had must stronger factor loadings than 377 ($\lambda_{young}$ = .616, $p$ < .001 and $\lambda_{old}$ = .549, $p$ < .001, compared to $\lambda_{young}$ = .391, $p$ < .001 and $\lambda_{old}$ = .384, $p$ < .001). Furthermore, both items had similar means ($M_{young}$ = 5.02 and $M_{old}$ = 5.28 for Item 377 and $M_{young}$ = 5.07 and $M_{old}$ = 5.30 for Item 380), and similar variability ($\sigma^2_{young}$ = 1.46 and $\sigma^2_{old}$ = 1.31 for item 377 and $\sigma^2_{young}$ = 1.47 and $\sigma^2_{old}$ = 1.39 for item 384). Thus Item 380 was retained over Item 377 at this step. Eleven Perceptual sensitivity items were dropped here according to Decision Rule 4.

A configural model was estimated with the remaining 10 items, all of which had significant standardized loadings greater than .40 in both groups, so no items were dropped based on Decision Rule 5 or Decision Rule 6. This configural model fit adequately, $\chi^2(70)$ = 128.01, $p$ < .0001, RMSEA = .064 (90% CI: .046 – .082), CFI = .937.

Next, all factor loadings were constrained to be equal across groups. This metric model fit comparatively well to the configural model, ΔLL $\chi^2(10)$ = 10.93, $p$ = .36, ΔRMSEA = -.003, ΔCFI = -.002. A scalar model was then estimated by constraining item intercepts across groups and freeing the factor mean for the older group. This scalar model fit comparatively worse than the metric model, ΔLL $\chi^2(9)$ = 63.21, $p$ < .0001, ΔRMSEA = .018, ΔCFI = -.055. Modification indices indicated scalar noninvariance for Item 339 (*"My child notices low-pitched noises such as the air-conditioner, heater, or* *refrigerator running or starting up."*), and Item 371 (*"My child seems to notice when s/he encounters small flaws in objects, such as a small stain, a chip in a plate, or a scratch on a counter."*). First, Item 339 was dropped and new models estimated, but the subsequent scalar model was still a comparatively poorer fit to the metric model, so Item 371 was dropped and a new set of models estimated. The scalar model still fit more poorly than the metric model, and modification indices indicated two problems with the remaining Item 354 (*"My child notices the smoothness or roughness of objects s/he touches."*): the item was scalar noninvariant and the item loading covaried with Item 342 (*"My child seems to notice when sounds change in volume, such as when someone turns the radio or television up or down."*). Because only eight items remained in the scale, Item 354 was dropped and exchanged with an item dropped earlier as its pair for a proposed covaried residual, Item 358 (*"My child likes to run his/her hand over things to see if they are smooth or rough."*). A new set of models were estimated, with the metric model fitting as well as the configural model, but the scalar model fit still fitting more poorly than the metric model. Modification indices indicated two item intercepts to be freed, Item 328 (*"My child will notice different aromas in nature (e.g., flowers), even when the scents are subtle."*) and Item 353 (*"My child seems to notice when a room is more warm or cool than expected."*). Only after freeing both of these intercepts did the scalar model fit as well as the metric model based on two of the three criteria, ΔLL $\chi^2(5) = 10.30$, $p = .07$, ΔRMSEA = .007, ΔCFI = -.008. The final scalar model with 8 items fit the data well, $\chi^2(53) = 64.60$, $p = .13$, RMSEA = .033 (90% CI: .000 - .058), CFI = .980.

For the original CBQ Perceptual Sensitivity scale, two items demonstrated configural noninvariance (Item 359 / CBQ 142 *"My child doesn't usually react to* *different textures of food."* and Item 387 / CBQ 84 *"My child doesn't usually comment on people's facial features, such as size of nose or mouth."*) and four demonstrated scalar noninvariance (Item 324 / CBQ 170 *"My child doesn't usually notice odors, such as perfume, smoke, cooking, etc."*, Item 338 / CBQ 105 *"My child usually comments if someone has an unusual voice."*, Item 363 / CBQ 98 *"My child is quickly aware of some new item in the living room."*, and Item 369 / CBQ 154 *"My child notices even little specks of dirt on objects."*). The final scalar invariance model contained 10 of the original 11 CBQ Perceptual Sensitivity scale items—the configurally noninvariant items were dropped—and fit poorly, $\chi^2(85) = 247.89$, $p < .001$, RMSEA = .098 (90% CI: .084 – .112), CFI = .755.

***Proposed Scales.*** The proposed scales based on statistical methods are presented in **Error! Reference source not found.**. Activation Control (8 items) contained two items in the cognitive activation content area (Items 21 & 25), three items from persistence (Items 30, 35, and 43), one item from physical activation (Item 56), and two items with unspecified content areas (Items 9 & 5). The content area of social activation was not represented in the Activation Control scale. Attention Shifting (Item 9 items) contained two items relating to dividing attention (Items 79 & 81), one orienting attention item (Item 86), three release of attention items (Items 94, 97, 106), and one shifting attention item (Items 111). Attentional Focusing (8 items) contained four concentration items (Items 125, 130, 131, & 141), three distractibility items (Items 143, 150, & 160), and one sustained attention item (Item 177). Inhibitory Control (10 items) contained two carefulness items (Items 191 & 194), four inhibiting behavior items (Items 218, 223, 240,

& R04), three waiting / delay of gratification items (Items 202, 207, & 216), and one

unspecified IC item (Item

Table 3.8

*Proposed revised CBQ Effortful Control scales based on statistical evaluation.*

| **Activation Control (8 items)** | | |
|---|---|---|
| **#** | **Content Area** | **Item** |
| 5 | *undefined* | My child follows rules for games, rather than making up their own rules to suit them. |
| 9 | *undefined* | When I'm busy, my child can find another activity to do when asked. |
| 21 | Cognitive Activation | My child likes to learn new skills, such as drawing or writing. |
| 25 | Cognitive Activation | My child will participate in a play activity, even if it is not the activity they would prefer. |
| 30 | Persistence | When picking up toys or doing other jobs, my child usually keeps at the task until it's done. |
| 35 | Persistence | My child begins but does not finish activities, such as crafts or puzzles. |
| 43 | Persistence | My child quickly gives up on an activity or project when s/he becomes bored of it. |
| 56 | Physical Activation | My child follows daily routines on his/her own, such as brushing his/her teeth. |
| **Attentional Focusing (8 items)** | | |
| **#** | **Content Area** | **Item** |
| 125 | Concentration | When trying to learn how a new toy works, my child concentrates intensely. |
| 130 | Concentration | When practicing an activity, my child has a hard time keeping her/his mind on it. |
| 131 | Concentration | My child has difficulty focusing on an activity or task, such as when listening to a story or working on a puzzle. |
| 141 | Concentration | My child pays close attention when something is being explained to him or her. |
| 143 | Distractibility | My child is easily distracted when listening to a story. |
| 150 | Distractibility | My child gets easily distracted when drawing, reading, or playing alone. |
| 160 | Distractibility | My child continues to focus on what they are doing, even if there are other things going on around them. |
| 177 | Sustained Attention | My child will maintain attention for a long time, even on activities that are not as pleasant/enjoyable (e.g., cleanup, homework). |

Table 3.8, continued

*Proposed revised CBQ Effortful Control scales based on statistical evaluation.*

**Attention Shifting (9 items)**

| # | Content Area | Item |
|---|---|---|
| 79 | Dividing Attention | My child sometimes doesn't seem to hear me when I talk to her/him. |
| 81 | Dividing Attention | My child can only pay attention to one activity or source of information at a time. |
| 86 | Orienting Attention | When asked to look at something, my child does so immediately. |
| 93 | Release of Attention | It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework. |
| 94 | Release of Attention | It is hard to get my child's attention when s/he is concentrating on something. |
| 97 | Release of Attention | When involved in an activity, my child answers quickly if I call him/her. |
| 99 | Release of Attention | My child has an easy time leaving play to come to dinner. |
| 106 | Release of Attention | My child has difficulty turning off the TV when told to do so. |
| 111 | Shifting Attention | If spoken to when watching TV, my child will reply appropriately then go back to watching TV. |

**Inhibitory Control (10 items)**

| # | Content Area | Item |
|---|---|---|
| 6 | *undefined* | My child has a hard time following instructions. |
| 191 | Carefulness | My child plays rough with their toys and damages them. |
| 194 | Carefulness | My child completes chores carefully, paying attention to details. |
| 202 | Waiting / DoG | My child has an easy time waiting to open a present. |
| 207 | Waiting / DoG | My child has a hard time waiting for his/her favorite snack or meal. |
| 216 | Waiting / DoG | When I’m on a phone call and my child wants to talk to me, s/he will wait until I’ve finished the call. |
| 218 | Inhibiting Behavior | My child can lower his/her voice when asked to do so. |
| 223 | Inhibiting Behavior | My child has trouble sitting still when s/he is told to (at movies, church, etc.). |
| 240 | Inhibiting Behavior | My child is usually able to resist temptation when told s/he is not supposed to do something. |
| R04 | Inhibiting Behavior | My child can easily stop a play activity when told to stop. |

Table 3.8, continued

*Proposed revised CBQ Effortful Control scales based on statistical evaluation.*

**Low Intensity Pleasure (9 items)**

| # | Content Area | Item |
|---|---|---|
| 253 | Cuddliness | My child seeks opportunities to get hugs and kisses from family members. |
| 255 | Cuddliness | My child enjoys sitting on a parent's lap. |
| 259 | Cuddliness | My child enjoys gentle rhythmic activities, such as rocking or swaying. |
| 273 | Listening | My child enjoys listening to music quietly. |
| 283 | Listening | My child enjoys just being talked to. |
| 285 | Listening | My child likes the sound of words, as in nursery rhymes or poems. |
| R06 | Listening | My child likes it when a parent or caregiver sings to him/her. |
| 310 | Quiet Activities | My child enjoys taking warm baths. |
| 318 | Quiet Activities | When around flowers, my child will take the time to smell them and enjoy their scent. |

**Perceptual Sensitivity (8 items)**

| # | Content Area | Item |
|---|---|---|
| 328 | Scent | My child will notice different aromas in nature (e.g., flowers), even when the scents are subtle |
| 331 | Sound | My child seems to listen to even quiet sounds. |
| 335 | Sound | My child notices when very quiet music is playing in the background. |
| 342 | Sound | My child seems to notice when sounds change in volume, such as when someone turns the radio or television up or down. |
| 353 | Temperature | My child seems to notice when a room is more warm or cool than expected. |
| 358 | Touch | My child likes to run his/her hand over things to see if they are smooth or rough. |
| 378 | Visual | My child comments when a parent has changed his/her appearance. |
| 380 | Visual | My child notices small changes in the environment, like lights getting brighter in a room. |

6). Low Intensity Pleasure (9 items) contained three items relating to cuddliness (Items 253, 255, & 259), four related to listening (Items 271, 283, 285, & R06), and two related to quiet activities (Items 310 & 318). Perceptual Sensitivity (8 items) contained one item relating to scent (Item 328), three relating to sound (Items 331, 335, & 342), one relating to temperature (Item 353), one relating to touch (Item 358), and two relating to visual sensitivity (Items 378 & 380).

**Theoretical Refinement**

I presented the proposed scales based on statistical considerations to content experts, along with a description of decision rules, scale definitions, and psychometric properties (factor loadings, CFA fit statistics, α, scale mean, scale standard deviation). The content experts made recommendations for changes on four scales: Activation Control, Attentional Focusing, Attention Shifting, and Inhibitory Control. When experts proposed dropping and replacing items, alternate items were considered based on content area coverage and the point during statistical evaluation at which the item was dropped. Items dropped based on Decision Rule 8 (scalar noninvariance) and Decision Rule 7 (metric noninvariance) were considered first as they typically demonstrated good factor loadings, and items dropped based on Decision Rule 4 (covaried residuals) were also considered if no remaining scale items were identified as their pair by modification indices. After dropping items, reintroduced items were added to the scale, and I conducted item analysis using the calibration sample, examining corrected item-total correlation and scale α to determine if the reintroduced item improved the scale. These modified scales were then presented to the content experts to approve or provide additional suggestions. Once scales had been finalized, a final round of invariance testing using the calibration sample was conducted to determine the degree to which scale items were measurement noninvariant. Items were not dropped during this process, rather, if an item demonstrated measurement noninvariance at some point, the corresponding parameter (loading or intercept) was freed between groups. Covaried residuals were also estimated if recommended by modification indices. Models were estimated until the metric model fit as well as the configural model, and the scalar model fit as well as the metric model.

**Recommendations and Modifications.** Based on content expert feedback, four of the proposed scales were modified on theoretical grounds: Activation Control, Attention Shifting, Attentional Focusing, and Inhibitory Control.

***Activation Control.*** Content experts noted that Activation Control seemed to be the most problematic scale. Three items were flagged for potentially not relating to overcoming an inclination or aversion: Item 21 (*"My child likes to learn new skills, such as drawing or writing."*, cognitive activation content area), Item 35 (*"My child begins but does not finish activities, such as crafts or puzzles."*, persistence), and Item 56 (*"My child follows daily routines on his/her own, such as brushing his/her teeth."*, physical activation). Five items were added in their place. Item 16 (*"My child takes care of needed tasks (e.g., washing hands, going to the bathroom, cleaning up) before playing."*) initially demonstrated covaried residuals with Item 56, but was considered a cognitive activation item, so this item was added to the scale. Item 33 (*"My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them."*, persistence) was initially dropped for displaying metric noninvariance and was added to the scale. Item 44 (*"When needing to walk long distances, my child will keep walking even when feeling tired."*, physical activation) initially demonstrated covaried residuals with an item that was subsequently dropped, and was added to the scale.

One content expert identified two Attention Shifting items as likely relating more closely to Activation Control because they involved overcoming an aversion or inclination. These two items (Item 93, *"It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework.",* and Item 99, *"My child has an easy time leaving play to come to dinner."*) were also added to the Activation Control scale, where they contributed considerably to the scale (improved α and mean inter-item correlation). One content expert also noted that *social activation* really overlaps with physical activation (e.g., they demand physically performing a task). Social activation may also overlap with aspects of shyness, another dimension of temperament related to extraversion / surgency. Given that no social activation items remained in the final scales and the perceived overlap with other constructs, the *social activation* content area was removed from the final scale description.

***Attention Shifting.*** The aforementioned two Attention Shifting items were removed from this scale and replaced with Item 85 (*"My child can focus attention promptly when cued, for example attending to directions."*), an orienting attention item that was dropped for demonstrating scalar noninvarience. Item 85 contributed well to the scale (improved α and mean inter-item correlation). An additional release of attention item that improved the scale was not found, and given that three other items representing this content area were already included in the scale, I felt it was unnecessary to include a replacement.

Dr. Mary Rothbart also recommended slightly altering the definition of Attention Shifting to include the word "voluntarily", (*"Capacity to <u>voluntarily</u> shift attention from one activity to another"*). Dr. Rothbart was not initially involved in revising scale and content area definitions and likely would have proposed this change during that stage. The other content experts either agreed or did not dissent to the change, and because this alteration also seems to better reflect the construct measured and scale items, this new wording was retained.

***Attentional Focusing.*** Two revised items were flagged by content experts. Item 131 (*"My child has difficulty focusing on an activity or task, such as when listening to a story or working on a puzzle."*, concentration) was flagged because an expert noted that this item overlaps with, and was in fact a rewrite of, Item 143 (CBQ 171 *"My child is easily distracted when listening to a story."*). Item 177 (*"My child will maintain attention for a long time, even on activities that are not as pleasant/enjoyable (e.g., cleanup, homework)."*, sustained attention) was also flagged because of overlap with Activation Control, as it refers to doing something unpleasant. These dropped items left the scale with three concentration items, three distractibility items, and no concentration items. In the interest of content balance, three alternate sustained attention items were reintroduced to the scale. Item 178 (*"My child will stay with the same activity for a half-hour or more if given the chance."*) and Item R01 (*"My child spends a long time engaged in drawing, coloring, or crafts."*) were initially dropped due to scalar noninvariance, but reintroduced at this point. Item 162 (*"My child sometimes becomes absorbed in a picture book and looks at it for a long time."*) was the last item to be dropped due to low loadings prior to invariance testing ($\lambda_{young}$ = .377, $\lambda_{old}$ = .289), and was reintroduced as well. All three items contributed well to the modified scale.

***Inhibitory Control.*** Two carefulness items were flagged by Dr. Rothbart for overlap with other constructs. Item 191 (*"My child plays rough with their toys and* *damages them."*) was deemed to clearly overlap with the temperament factor Surgency (e.g., high intensity pleasure, activity level, impulsivity, etc.; Rothbart et al., 2001). Item 194 (*"My child completes chores carefully, paying attention to details."*) was deemed to overlap with the temperament construct of Fear. Dr. Rothbart noted that "[w]hen one takes care, one tries to avoid a mistake, and an overlap with fear is present" (M. Rothbart, personal communication, May 14, 2017). She went on to note that caution is therefore a component of fear, suggesting that in hindsight, caution and carefulness should not have been included in the original CBQ Inhibitory Control scales. The other content experts either agreed to this point, deferred to Dr. Rothbart's judgment, or provided no dissent. This resulted in the two carefulness items being dropped from the scale, and the scale definition of Inhibitory Control to be revised as *"The capacity to suppress or moderate desired behaviors and delay actions."* No additional items were added to the scale as replacements.

***Low Intensity Pleasure.*** One proposed modification to the Low Intensity Pleasure scale involved Item 253 (*"My child seeks opportunities to get hugs and kisses from family members."*). It was suggested that Item 253 may refer to incentive motivation as opposed to enjoyment—which is how Low Intensity Pleasure is defined—and therefore *"seeks opportunities to get"* could be reworded in the future, perhaps as *"takes pleasure in getting"*. Because no other problems were identified for scale modification, the scale remained unchanged.

## Final Scales

I presented the modified scales to the content experts, and they gave their approval. The modified Activation Control scale contains 10 items, including two cognitive activation items, three persistence items, one physical activation items, two unspecified AC items, and two misplaced AS items. The modified Attention Shifting scale contains eight items, including two dividing attention items, two orienting attention items, three release of attention items, and one shifting attention items. The modified Attentional Focusing scale contains three concentration items, three distractibility items, and three sustained attention items. The modified Inhibitory Control scale contains eight items, consisting of four behavior inhibition items, three waiting / delay of gratification items, and one item of an unspecified content area. The unchanged Low Intensity pleasure scale contains nine items, and the Perceptual Sensitivity scale contains 8 items. These final scales are presented in

Table 3.9

*Final revised CBQ Effortful Control scales.*

, and the final scale and content area definitions are presented in Table 3.10. I have also created a printable questionnaire with randomized CBQ-R EC questions (Appendix A), and a corresponding score sheet with scale definitions and psychometric properties (Appendix B).

**Invariance testing modified scales.** The four modified scales were then subjected to measurement invariance testing. Only the final model structures are described here, in lieu of detailing the entire model-fitting process.

***Activation Control***. Results of invariance testing produced an Activation Control scalar model with a few structural modifications. Residuals of Item 33 (*"My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them."*) and Item 43 (*"My child quickly gives up on an activity or project when s/he becomes bored of it."*) were covaried resulting in a standardized error covariance = -.024, $p$ = .780 for the younger group, and a standardized error covariance = .285, $p$ = .001 for the older group. Residuals of Item 93 (*"It takes a*

Table 3.9

*Final revised CBQ Effortful Control scales.*

| **Activation Control (10 items)** | | |
|---|---|---|
| **#** | **Content Area** | **Item** |
| 16 | Cognitive Activation | My child takes care of needed tasks (e.g., washing hands, going to the bathroom, cleaning up) before playing. |
| 25 | Cognitive Activation | My child will participate in a play activity, even if it is not the activity they would prefer. |
| 30 | Persistence | When picking up toys or doing other jobs, my child usually keeps at the task until it's done. |
| 33 | Persistence | My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them. |
| 43 | Persistence | My child quickly gives up on an activity or project when s/he becomes bored of it. |
| 44 | Physical Activation | When needing to walk long distances, my child will keep walking even when feeling tired. |
| 5 | *unspecified* | My child follows rules for games, rather than making up their own rules to suit them. |
| 9 | *unspecified* | When I'm busy, my child can find another activity to do when asked. |
| 93 | *unspecified* | It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework. |
| 99 | *unspecified* | My child has an easy time leaving play to come to dinner. |
| **Attentional Focusing (9 items)** | | |
| **#** | **Content Area** | **Item** |
| 125 | Concentration | When trying to learn how a new toy works, my child concentrates intensely. |
| 130 | Concentration | When practicing an activity, my child has a hard time keeping her/his mind on it. |
| 141 | Concentration | My child pays close attention when something is being explained to him or her. |
| 143 | Distractibility | My child is easily distracted when listening to a story. |
| 150 | Distractibility | My child gets easily distracted when drawing, reading, or playing alone. |
| 160 | Distractibility | My child continues to focus on what they are doing, even if there are other things going on around them. |
| 162 | Sustained Attention | My child sometimes becomes absorbed in a picture book and looks at it for a long time. |
| 178 | Sustained Attention | My child will stay with the same activity for a half-hour or more if given the chance. |
| R01 | Sustained Attention | My child spends a long time engaged in drawing, coloring, or crafts. |

Table 3.9, continued

*Final revised CBQ Effortful Control scales.*

| **Attention Shifting (8 items)** | | |
|---|---|---|
| **#** | **Content Area** | **Item** |
| 79 | Dividing Attention | My child sometimes doesn't seem to hear me when I talk to her/him. |
| 81 | Dividing Attention | My child can only pay attention to one activity or source of information at a time. |
| 86 | Orienting Attention | When asked to look at something, my child does so immediately. |
| 85 | Orienting Attention | My child can focus attention promptly when cued, for example attending to directions. |
| 94 | Release of Attention | It is hard to get my child's attention when s/he is concentrating on something. |
| 97 | Release of Attention | When involved in an activity, my child answers quickly if I call him/her. |
| 106 | Release of Attention | My child has difficulty turning off the TV when told to do so. |
| 111 | Shifting Attention | If spoken to when watching TV, my child will reply appropriately then go back to watching TV. |
| **Inhibitory Control (8 items)** | | |
| **#** | **Content Area** | **Item** |
| 218 | Inhibiting Behavior | My child can lower his/her voice when asked to do so. |
| 223 | Inhibiting Behavior | My child has trouble sitting still when s/he is told to (at movies, church, etc.). |
| 240 | Inhibiting Behavior | My child is usually able to resist temptation when told s/he is not supposed to do something. |
| R04 | Inhibiting Behavior | My child can easily stop a play activity when told to stop. |
| 6 | *unspecified* | My child has a hard time following instructions. |
| 202 | Waiting / DoG | My child has an easy time waiting to open a present. |
| 207 | Waiting / DoG | My child has a hard time waiting for his/her favorite snack or meal. |
| 216 | Waiting / DoG | When I’m on a phone call and my child wants to talk to me, s/he will wait until I’ve finished the call. |
| **Low Intensity Pleasure (9 items)** | | |
| **#** | **Content Area** | **Item** |
| 253 | Cuddliness | My child seeks opportunities to get hugs and kisses from family members. |
| 255 | Cuddliness | My child enjoys sitting on a parent's lap. |

Table 3.9, continued

*Final revised CBQ Effortful Control scales.*

| | | |
|---|---|---|
| 259 | Cuddliness | My child enjoys gentle rhythmic activities, such as rocking or swaying. |
| 273 | Listening | My child enjoys listening to music quietly. |
| 283 | Listening | My child enjoys just being talked to. |
| 285 | Listening | My child likes the sound of words, as in nursery rhymes or poems. |
| R06 | Listening | My child likes it when a parent or caregiver sings to him/her. |
| 310 | Quiet Activities | My child enjoys taking warm baths. |
| 318 | Quiet Activities | When around flowers, my child will take the time to smell them and enjoy their scent. |

**Perceptual Sensitivity (8 items)**

| **#** | **Content Area** | **Item** |
|---|---|---|
| 328 | Scent | My child will notice different aromas in nature (e.g., flowers), even when the scents are subtle. |
| 331 | Sound | My child seems to listen to even quiet sounds. |
| 335 | Sound | My child notices when very quiet music is playing in the background. |
| 342 | Sound | My child seems to notice when sounds change in volume, such as when someone turns the radio or television up or down. |
| 353 | Temperature | My child seems to notice when a room is more warm or cool than expected. |
| 358 | Touch | My child likes to run his/her hand over things to see if they are smooth or rough. |
| 378 | Visual | My child comments when a parent has changed his/her appearance. |
| 380 | Visual | My child notices small changes in the environment, like lights getting brighter in a room. |

Table 3.10

*Final revised CBQ Effortful Control scale and content area definitions.*

**Activation Control:** The capacity to perform a subdominant action against a natural inclination or aversion.

1. Cognitive activation: the ability to engage in cognitively demanding subdominant actions
2. Physical activation: the ability to engage in physically demanding subdominant actions
3. Persistence: maintaining attentional focus until task completion

**Attentional Focusing:** Capacity to maintain attentional focus upon task-related channels.

1. Concentration: the intensity of attentional focus
2. Distractibility: the ease with which attention is involuntarily broken
3. Sustained Attention: the duration of attentional focus

**Attention Shifting:** Capacity to voluntarily shift attention from one activity to another.

1. Dividing attention: the ability to attend to competing stimuli simultaneously
2. Release of attention: the ease with which attention is voluntarily broken
3. Shifting attention: the ease with which attention is shifted between stimuli or tasks
4. Orienting Attention: the ease with which attentional focus is initiated, including the vigilance of being alert to environmental stimuli

**Inhibitory Control:** The capacity to suppress or moderate desired behaviors and delay actions.

1. Inhibiting behavior: the ability to suppress a desired behavior, including inhibiting a response, slowing motor actions, and lowering voice
2. Waiting: the ability to delay an action, either in the presence or absence of future gain or reward

**Low Intensity Pleasure:** Amount of pleasure or enjoyment related to situations involving low stimulus intensity, rate, complexity, novelty, and incongruity.

1. Enjoyment from quiet activities: the amount of enjoyment derived from simple, quiet activities.
2. Enjoyment from listening: the amount of enjoyment derived from listening to sounds, words, or song
3. Cuddliness: the amount of enjoyment derived from gentle contact and movement

**Perceptual Sensitivity:** Amount of detection of slight, low intensity stimuli from the external environment.

1. Visual sensitivity: noticing changes in the environment, including new objects, small details, and changes in another’s physical appearance
2. Sensory sensitivity: detecting subtle details in non-visual sensations, such as quiet sounds, and new or unique scents, flavors, and textures

*long time to get my child engaged in an activity they are not interested in, such as chores or homework."*) and Item 30 (*"When picking up toys or doing other jobs, my child usually keeps at the task until it's done."*) were also covaried, resulting in a standardized error covariance = .027, $p$ = .758 for the younger group, and an error covariance = .280, $p$ = .002 for the older group. All items demonstrated metric invariance. Item 93 demonstrated scalar noninvariance, as did Item 5 (*"My child follows rules for games, rather than making up their own rules to suit them."*), Item 9 (*"When I'm busy, my child can find another activity to do when asked."*), and Item 44 (*"When needing to walk long distances, my child will keep walking even when feeling tired."*). The remaining six items were scalar invariant.

***Attentional Focusing.*** Residuals of Item 141 (*"My child pays close attention when something is being explained to him or her."*) was covaried with Item 125 (*"When trying to learn how a new toy works, my child concentrates intensely."*) and Item 143 (*"My child is easily distracted when listening to a story."*). Standardized error covariance between Item 141 and Item 125 = .223, $p$ = .009 for the younger group, and = .125, $p$ = .062 for the older group. Standardized error covariance between Item 141 and Item 143 = .060, $p$ = .440 for the younger group, and = .269, $p$ = .001 for the older group. Residuals for Item R01 (*"My child spends a long time engaged in drawing, coloring, or crafts."*) and Item 162 (*"My child sometimes becomes absorbed in a picture book and looks at it for a long time."*) were also covaried, so that the standardized error covariance = .268, $p$ < .001 for the younger group, and = .062, $p$ = .446 for the older group. All items demonstrated metric invariance. Item R01 and Item 178 (*"My child will stay with the* *same activity for a half-hour or more if given the chance."*) demonstrated scalar noninvariance. All seven other items were scalar invariant.

***Attention Shifting.*** No items required covaried residual terms. All items demonstrated metric invariance. The newly introduced item, Item 85 (*"My child can focus attention promptly when cued, for example attending to directions."*) displayed scalar noninvariance. All seven other items were scalar invariant.

***Inhibitory Control***. No items required covaried residual terms. All items demonstrated both metric and scalar invariance.

## Step 6: Compare Original and Revised Scales

### Analysis Procedure

The validation sample was used to evaluate and compare the original and revised EC scales in terms of reliability and validity evidence. To be consistent with the approach taken in Study 1, the scales are also evaluated by examining response frequency (low use: < 5%) using the validation sample, and independent samples t-tests comparing younger and older age groups using the calibration sample.

**Reliability Evidence.** Internal consistency reliability will be evaluated using standardized Cronbach's α and McDonald's Omega. Whereas alpha holds the usually untenable assumption of tau equivalency—that the covariance between an item and the true score are equal across all items (DeVellis, 2017; Dunn, Baguley, & Brunsden, 2014)—coefficient Omega does not. Thus, Omega is a more practical alternative to alpha, especially as an index of internal consistency within a CFA framework, albeit not as ubiquitous. Based on DeVellis (2017), α values will be evaluated as follows: values below .60 are unacceptable; values .60 to .65 are undesirable; values between .65 and .70

are minimally acceptable / mediocre; values .70 to .80 are respectable; values between .80 and .90 are very good; and values .90 and above indicate the scale could be shortened (p. 145). I could find no guidelines for interpreting Omega, but as it is on a similar scale to alpha (i.e., 0.0 – 1.0) and produces similar values, I will use the same guidelines for interpretation.

Temporal stability (i.e., test-retest reliability) will be evaluated using a subsample of participants by correlating composite scores from initial participation with composite scores from responding to the same items again one month later.

**Validity Evidence.** Content-related evidence of validity has already been provided by content experts, who had final say on the revised scales. Furthermore, content-related evidence of validity will be demonstrated by content overlap between original and revised scales. Construct-related evidence of validity will be established by correlating the composite score of the revised scales with the composite scores of the original scales. Although a correlation of 1.0 would indicate a revised scale is indistinguishable from the original scale in terms of scores, large correlations less than 1.0 would suggest the scales have a strong linear relationship and that, given content overlap, they are measuring similar constructs. Additional construct-related evidence of validity will be presented by estimating a single-group unstructured CFA (with no covaried residual item variances) and evaluating model fit with (a) $\chi^2$ test of model fit, with non-significant values indicating good fit, (b) RMSEA values less than .06 indicating “close” fit , values less than .08 indicating “acceptable” fit, values between .08 and .10 indicating “mediocre” fit, and values greater than .10 as “unacceptable”, and (c) CFI values greater than .90 indicating acceptable fit, and CFI values greater than .95

indicating good fit (Brown, 2006; Hu & Bentler, 1998). Regarding significance values, the $\chi^2$ statistic is inflated by sample size, with larger samples more likely to be significant. The convention of $\alpha$ = .05 for $\chi^2$ tests may be too stringent (Brown, 2006) and therefore for our purposes $\chi^2$ values significant below $\alpha$ = .01 will be considered "poor fit", and non-significant values will be considered good fit.

Internal structure validity evidence (i.e. dimensionality; Sireci & Sukin, 2013) will be evaluated in a Principal Components Analysis (PCA) context and consulting multiple rules to determine the number of components present. Raîche, Riopel, and Blais (2006) propose four tests of PCA to be considered for determining the number of principal components, based on the Cattell Scree test (Cattell, 1966). The Kaiser-Guttman rule (Guttman, 1954; Kaiser, 1960) proposes that a factorized correlation matrix represents only the number of components with eigenvalues greater than 1.00. Parallel analysis is a modification of the Kaiser-Guttman rule, based on the proportion of eigenvalues obtained from a set of correlation matrices with random observations of p variables, where p is equal to the number of variables or items in question (Horn, 1965). Optimal coordinates places a line of predicted eigenvalues based on observed eigenvalues, and the number of principal components is equal to the number of eigenvalues greater than that predicted for each component (Raîche, Riopel, & Blais, 2006). Finally, the Scree Test Acceleration Factor identifies the number of principal based on the point in the scree plot at which the slopes of lines connecting eigenvalues changes abruptly (Raîche, Riopel, & Blais, 2006).

**Participants**

Although the calibration sample excluded responses for 5 year-old children, the validation sample included data for these children to ensure the scale evaluation was based on a sample for the entire intended age-range. The validation sample consisted of 272 parents of children ages 3-7 (143 female girls, mean age = 62.97 months, SD = 16.10) with N = 100 parents of 3-4 year olds (54 female children, mean age = 46.03, SD = 8.75), N = 93 parents of 5 year olds (42 female children, mean age= 65.31, SD = 3.95), and N= 79 parents of 6-7 year olds (47 female children, mean age = 81. 65, SD = 7.54). Parents were overwhelmingly female (N = 253, mean age = 34.79 years, SD = 4.96) and white (89.0%), with some parents reporting as mixed race (4.0%), Asian (2.2%), Hispanic (1.1%), Black (0.7%), other (0.7%), or refused to report (2.2%). Children were also overwhelmingly white (82.4%), with others reported as mixed race (10.3%), Other (1.8%), Asian (1.1%), Hispanic or Black (0.7% each), or not reported (2.9%). Most parents were married (83.8%), with the remaining parents either never married (6.3%), separated or divorced (4.8%), some other marital status (4.0%), or refused to answer (1.1%). Parents reported having about two children in the household (M = 2.05, SD = .85) and a median household income of between \$80,000 and \$89,999 per year. More than twice as many parents took questionnaire Version 2 (70.2%) than Version 1 (29.8%). Most parents reported residing in the United States (81.3%), with others living in the United Kingdom or Canada (5.5% each), Australia (4.0%), South Africa (1.5%), or elsewhere (2.2%). Based on ZIP Codes, the U.S. residents lived primarily in Urbanized Areas (80.6%) with the rest living in Urban Clusters (10.6%) or Rural Areas (8.8%).

Of the 536 parents who indicated a willingness to retake the questionnaire, 150 were contacted via email one month after having initially taken the questionnaire, and were asked to complete the questionnaire again. One third of these parents (N=50) began the retake questionnaire. Parents who completed the retake questionnaire (N=39, 38 women, mean age = 33.69 years, SD = 4.29) were white (87.2%), mixed race (10.3%) or reported another ethnicity (2.6%). Median household income was $90,000 - $99,999 per year. Children reported on ($M_{age}$ = 59.46 months, SD = 15.48) were white (82.1%), mixed race (15.4%), or other (2.6%). Twenty-one children were male (54%), 17 were female (44%), and 1 was reported as transgender (3%). Families who resided in the United States (71.8%) came from Urbanized Areas (89.3%) or Urban Clusters (10.7%). The rest of the retake sample resided in the United Kingdom (10.3%), Australia (7.7%), Canada (5.1%), New Zealand and South Africa (2.6% each).

## Results

**Reliability Evidence.** Scale reliability was computed using R version 3.4.0 (R Core Team, 2017) with the “omega” function from R package “psych” version 1.7.3 (Revelle, 2017). Table 3.11 presents the classical test theory measures of original and revised scale performance. Omega values were estimated based on a single-factor model. Standardized α values are reported here, denoted as “Std α”.

The revised Attentional Focusing scale has the same number of items as the original scale (9), but a higher internal consistency (α = .796 versus α = .748; omega = .800 versus omega .767), higher mean inter-item correlation (*r* = .302 versus *r* =.245), a slightly higher mean (*M* = 4.907 vs. *M* = 4.774), and slightly more variability (*SD* = 0.770 versus *SD* = 0.734).

Table 3.11

*Classical Test Theory evaluation of original and revised CBQ Effortful Control scales.*

| | Items | α | Omega | Mean *r* | Scale Mean | Scale SD |
|---|---|---|---|---|---|---|
| Activation Control - Revised | 10 | .819 | .823 | .312 | 4.441 | 0.852 |
| Attentional Focusing - Original | 9 | .748 | .767 | .245 | 4.774 | 0.734 |
| Attentional Focusing - Revised | 9 | .796 | .800 | .302 | 4.907 | 0.770 |
| Attention Shifting - Original | 5 | .821 | .825 | .478 | 4.567 | 0.964 |
| Attention Shifting - Revised | 8 | .835 | .839 | .389 | 4.649 | 0.909 |
| Inhibitory Control - Original | 13 | .877 | .882 | .353 | 4.876 | 0.846 |
| Inhibitory Control - Revised | 8 | .847 | .849 | .410 | 4.392 | 0.949 |
| Low Intensity Pleasure - Original | 13 | .769 | .775 | .203 | 5.543 | 0.631 |
| Low Intensity Pleasure - Revised | 9 | .731 | .734 | .230 | 5.476 | 0.765 |
| Perceptual Sensitivity - Original | 12 | .779 | .790 | .226 | 4.794 | 0.815 |
| Perceptual Sensitivity - Revised | 8 | .820 | .824 | .362 | 4.714 | 1.020 |
| All Items - Original | 52 | .916 | .919 | .169 | 4.977 | 0.566 |
| All Items - Revised | 52 | .920 | .913 | .209 | 4.767 | 0.642 |

Note: Validation Sample. α = Standardized Cronbach's α, Mean *r* = mean inter-item bivariate correlation.

The revised Attention Shifting scale contains three more item than the original scale (8 versus 5), has slightly higher internal consistency ($\alpha$ = .835 versus $\alpha$ = .821; omega = .839 versus omega = .825), lower mean inter-item correlation ($r$ = .389 versus $r$ = .478), slightly higher mean ($M$ = 4.649 vs. $M$ = 4.567), and slightly less variability ($SD$ = .909 versus $SD$ = .964).

The revised Inhibitory Control scale has five fewer items than the original (8 versus 13), slightly lower internal consistency ($\alpha$ = .847 versus $\alpha$ = .877; omega = .849

versus .882), higher mean inter-item correlation (*r* = .410 versus *r* = .353), a lower scale mean (*M* = 4.392 versus *M* = 4.876), and more variability (*SD* = .949 versus *SD* = .846).

The revised Low Intensity Pleasure scale has four fewer items than the original (9 versus 13), slightly lower internal consistency (*α* = .731 versus *α* = .769; omega = .734 versus .775), higher mean inter-item correlation (*r* = .230 versus *r* = .203), a slightly lower scale mean (*M* = 5.476 versus *M* = 5.543), and more variability (*SD* = .765 versus *SD* = .631).

The revised Perceptual Sensitivity scale has four fewer items than the original (8 versus 12), higher internal consistency (*α* = .820 versus *α* = .779; omega = .824 versus omega = .790), higher mean inter-item correlation (*r* = .362 versus *r* = .226), a lower scale mean (*M* = 4.714 versus *M* = 4.794), and more variability (*SD* = 1.020 0versus *SD* = .815).

The revised Activation Control scale contained 10 items, with very good internal consistency (*α* = .819; omega = .823), mean inter-item correlation of *r* = .312, scale mean closer to 4.0 than most of the other revised scales (*M* = 4.441) and similar variability to the other scales (*SD* = .852).

Overall, the six revised scales have the same number of items as the five original scales (52), with similar internal consistency (*α* = .920 versus *α* = .916; omega = .913 versus .919), higher inter-item correlation (*r* = .209 versus *r* = .169), a lower scale mean (*M* = 4.767 versus *M* = 4.977), and greater variability (*SD* = .642 versus *SD* = .566).

Based on DeVellis’ interpretation, two of the revised scales had *respectable* internal consistency (Attentional Focusing and Low Intensity Pleasure) and four had *very good* internal consistency (Activation Control, Attention Shifting, Inhibitory Control, and

Perceptual Sensitivity). Comparatively, three of the five original scales showed *respectable* internal consistency (Attentional Focusing, Low Intensity Pleasure, and Perceptual Sensitivity), whereas only two original scales showed *very good* internal consistency (Attention Shifting and Inhibitory Control).

Temporal Stability is presented in Table 3.12. All scales had test-retest Pearson correlations higher than $r$ = .75. Original scales had slightly higher temporal stability than the revised scales for Attentional Focusing ($r$ = .90 versus $r$ = .82), Inhibitory Control ($r$ = .82 versus $r$ = .76), Low Intensity Pleasure ($r$ = .82 versus $r$ = .79), and Perceptual Sensitivity ($r$ = .79 versus $r$ = .78). The revised Attention Shifting scale had slightly higher temporal stability than the original ($r$ = .81 versus $r$ = .76), and the new Activation Control scale had comparable temporal stability to the other scales ($r$ = .82).

Table 3.12

*Temporal stability for original and revised CBQ Effortful Control scales.*

| Scale | Original | Revised |
|---|---|---|
| Activation Control | - | .82 |
| Attentional Focusing | .90 | .82 |
| Attention Shifting | .71 | .81 |
| Inhibitory Control | .82 | .76 |
| Low Intensity Pleasure | .82 | .81 |
| Perceptual Sensitivity | .79 | .78 |
| All Items | .89 | .86 |

$N$ = 39. All values are significant at $p < .01$. Coefficients are bivariate correlations (Pearson's $r$) of one-month test-retest reliability.

**Validity Evidence.** ***Inter-Scale Correlations.*** Table 3.13 presents correlations between original and revised scales from the validation sample. All revised scales correlated highly with the corresponding original scale (Attentional Focusing $r$ = .88; Attention Shifting $r$ = .88 Inhibitory Control $r$ = .88; Low Intensity Pleasure $r$ = .86;

Table 3.13

*Bivariate correlations between original and revised CBQ EC scales.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Attentional Focusing - Original | | | | | | | | | | | | |
| 2 Attention Shifting - Original | .49 | | | | | | | | | | | |
| 3 Inhibitory Control - Original | .64 | .71 | | | | | | | | | | |
| 4 Low Intensity Pleasure - Original | .48 | .37 | .43 | | | | | | | | | |
| 5 Perceptual Sensitivity - Original | .27 | .12 | .26 | .35 | | | | | | | | |
| 6 All Items - Original | .77 | .69 | .84 | .72 | .61 | | | | | | | |
| 7 Activation Control - Revised | .72 | .77 | .76 | .44 | .16 | .75 | | | | | | |
| 8 Attentional Focusing - Revised | **.88** | .49 | .61 | .51 | .30 | .75 | .65 | | | | | |
| 9 Attention Shifting - Revised | .55 | **.90** | .73 | .41 | .19 | .72 | .76 | .56 | | | | |
| 10 Inhibitory Control - Revised | .60 | .74 | **.88** | .32 | .15 | .72 | .78 | .57 | .75 | | | |
| 11 Low Intensity Pleasure - Revised | .41 | .32 | .37 | **.86** | .36 | .64 | .40 | .45 | .36 | .26 | | |
| 12 Perceptual Sensitivity - Revised | .29 | .09[a] | .23 | .39 | **.80** | .54 | .16 | .34 | .17 | .12[a] | .42 | |
| 13 All Items - Revised | .78 | .75 | .81 | .65 | .45 | **.93** | .86 | .80 | .82 | .79 | .64 | .51 |

[a] Not significant at $p < .05$. All other values are significant at $p < .05$.

**Bolded** values are correlations between original and revised scales.

Note: Validation sample

Perceptual Sensitivity $r = .80$; All items $r = .93$), providing some evidence of construct validity.

***Confirmatory Factor Analysis.*** To provide further evidence of construct validity, unconstrained (no covaried residuals or constrained parameters) CFA models were estimated in Mplus 7.11 (Muthén & Muthén, 1998-2013) with robust maximum likelihood estimation (MLR) using the validation sample for both original and revised scales.

The new Activation Control unconstrained CFA model (Table 3.14) had standardized factor loadings ranging from $\lambda = .427$ to $\lambda = .742$, and demonstrated acceptable fit, $\chi^2(35) = 76.90$, $p < .0001$, RMSEA = .066 (90% CI: .046 - .086), CFI = .925.

Original and revised Attentional Focusing unconstrained CFA models are presented in Table 3.15. The revised model had standardized factor loadings ranging from $\lambda =.312$ to $\lambda = .721$, and demonstrated acceptable fit, $\chi^2(27) = 42.67$, $p < .0001$, RMSEA = .079 (90% CI: .057 - .101), CFI = .903. The original model had considerably lower standardized factor loadings ranging from $\lambda = -.019$ to $\lambda = .769$, with two loadings lower than the recommended threshold of .300 for meaningful contribution to the factor (Item 41 *"My child has difficulty leaving a project s/he has begun."*, $\lambda = -.019$, $p = .802$; and Item 162 *"My child sometimes becomes absorbed in a picture book and looks at it for a long time."*, $\lambda = .284$, $p < .001$). The CFA for the original items demonstrated poor fit, $\chi^2(27) = 83.74$, $p < .0001$, RMSEA = .088 (90% CI: .067 - .110), CFI = .874.

Original and revised Attention Shifting unconstrained CFA models are presented in Table 3.16. The revised model had standardized factor loadings ranging from $\lambda = .510$

Table 3.14

*Activation Control unconstrained CFA models.*

| Standardized Factor Loadings | Revised |
|---|---|
| 5. My child follows rules for games, rather than making up their own rules to suit them. | .476 |
| 9. When I'm busy, my child can find another activity to do when asked. | .600 |
| 16. My child takes care of needed tasks (e.g., washing hands, going to the bathroom, cleaning up) before playing. | .653 |
| 25. My child will participate in a play activity, even if it is not the activity they would prefer. | .427 |
| 30. When picking up toys or doing other jobs, my child usually keeps at the task until it's done. | .742 |
| 33. My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them. | .452 |
| 43. My child quickly gives up on an activity or project when s/he becomes bored of it. | .511 |
| 44. When needing to walk long distances, my child will keep walking even when feeling tired. | .450 |
| 93. It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework. | .738 |
| 99. My child has an easy time leaving play to come to dinner. | .529 |
| Model Fit | |
| Chi-Square | 76.90 |
| df | 35 |
| p | < .0001 |
| RMSEA | .066 |
| 90% CI | .046 - .086 |
| CFI | .925 |

Table 3.15

*Attentional Focusing unconstrained CFA models.*

| Items | Standardized Loadings | |
|---|---|---|
| | Original | Revised |
| 30. When picking up toys or doing other jobs, my child usually keeps at the task until it's done. | .610 | |
| 40r. My child will move from one task to another without completing any of them. | .756 | |
| 41. My child has difficulty leaving a project s/he has begun. | -.019[a] | |
| 127. When drawing or coloring in a book, my child shows strong concentration. | .443 | |
| 146r. My child has a hard time concentrating on an activity when there are distracting noises. | .458 | |
| 165. My child can spend a long time engaged in play with toys, drawing, or coloring. | .554 | |
| **130r. When practicing an activity, my child has a hard time keeping her/his mind on it.** | **.769** | **.721** |
| **143r. My child is easily distracted when listening to a story.** | **.639** | **.597** |
| **162. My child sometimes becomes absorbed in a picture book and looks at it for a long time.** | **.284[a]** | **.312** |
| 125. When trying to learn how a new toy works, my child concentrates intensely. | | .412 |
| 141. My child pays close attention when something is being explained to him or her. | | .594 |
| 150. My child gets easily distracted when drawing, reading, or playing alone. | | .675 |
| 160. My child continues to focus on what they are doing, even if there are other things going on around them. | | .504 |
| 178. My child will stay with the same activity for a half-hour or more if given the chance. | | .583 |
| R01. My child spends a long time engaged in drawing, coloring, or crafts. | | .541 |
| Model Fit | | |
| Chi-Square | 83.74 | 42.67 |
| *df* | 27 | 27 |
| *p* | < .0001 | < .0001 |
| RMSEA | .088 | .079 |
| 90% CI | .067 - .110 | .057 - .101 |
| CFI | .874 | .903 |

[a] Non-significant, $p > .05$.

**Bolded** items appear on both versions.

Table 3.16

*Attention Shifting unconstrained CFA models.*

| | Standardized Loadings | |
|---|---|---|
| Items | Original | Revised |
| 99. My child has an easy time leaving play to come to dinner. | .559 | |
| 100r. My child has a lot of trouble stopping an activity when called to do something else. | .823 | |
| 115. My child can easily shift from one activity to another. | .708 | |
| **79r. My child sometimes doesn't seem to hear me when I talk to her/him.** | **.715** | **.786** |
| **94r. It is hard to get my child's attention when s/he is concentrating on something.** | **.666** | **.702** |
| 81. My child can only pay attention to one activity or source of information at a time. | | .510 |
| 85. My child can focus attention promptly when cued, for example attending to directions. | | .664 |
| 86. When asked to look at something, my child does so immediately. | | .559 |
| 97. When involved in an activity, my child answers quickly if I call him/her. | | .691 |
| 106. My child has difficulty turning off the TV when told to do so. | | .520 |
| 111. If spoken to when watching TV, my child will reply appropriately then go back to watching TV. | | .746 |
| Model Fit | | |
| Chi-Square | 13.88 | 20.79 |
| *df* | 5 | 20 |
| *p* | .016 | .410 |
| RMSEA | .081 | .012 |
| 90% CI | .032 - .133 | .000 - .054 |
| CFI | .975 | .999 |

**Bolded** items appear on both versions.

to $\lambda$ = .786, and demonstrated very good fit, $\chi^2(20) = 20.79$, $p$ = .410, RMSEA = .012 (90% CI: .000 - .054), CFI = .999. The original model had a slightly higher range of standardized factor loadings ranging from $\lambda$ = .559 to $\lambda$ = .823, and demonstrated acceptable fit, $\chi^2(5) = 13.88$, $p$ = .016, RMSEA = .081 (90% CI: .032 - .133), CFI = .975.

Original and revised Inhibitory Control unconstrained CFA models are presented in Table 3.17. The revised model had standardized factor loadings ranging from $\lambda$ = .517 to $\lambda$ = .764, and demonstrated very good fit, $\chi^2(20) = 25.09$, $p$ = .198, RMSEA = .031 (90% CI: .000 - .064), CFI = .991. The original model had somewhat lower standardized factor loadings ranging from $\lambda$ = .329 to $\lambda$ = .780, and demonstrated acceptable fit, $\chi^2(35)$ = 76.90, $p$ < .0001, RMSEA = .066 (90% CI: .046 - .086), CFI = .925.

Original and revised Low Intensity Pleasure unconstrained CFA models are presented in Table 3.18. The revised model had standardized factor loadings ranging from $\lambda$ = .388 to $\lambda$ = .712, and demonstrated acceptable fit, $\chi^2(27) = 49.93$, $p$ = .005, RMSEA = .056 (90% CI: .031 - .080), CFI = .917. The original model had somewhat lower standardized factor loadings ranging from $\lambda$ = .210 to $\lambda$ = .696, and two items with loadings less than the recommended threshold of .300 (Item 282 *"My child rarely enjoys just being talked to."*, $\lambda$ = .219, $p$ = .005; and Item 303 *"My child doesn't care much for quiet games."*, $\lambda$ = .210, $p$ = .003). The original scale CFA demonstrated mediocre fit, $\chi^2(65) = 172.71$, $p$ < .0001, RMSEA = .078 (90% CI: .064 - .092), CFI = .780.

Original and revised Perceptual Sensitivity unconstrained CFA models are presented in Table 3.19. The revised model had standardized factor loadings ranging from $\lambda$ = .402 to $\lambda$ = .767, and demonstrated acceptable fit, $\chi^2(20) = 44.10$, $p$ = .002, RMSEA = .067 (90% CI: .040 - .093), CFI = .947. The original model had somewhat

Table 3.17

*Inhibitory Control unconstrained CFA model.*

| Items | Standardized Loadings | |
|---|---|---|
| | Original | Revised |
| 185. My child is not very careful and cautious in crossing streets. | .329 | |
| 192. My child approaches places s/he has been told are dangerous slowly and cautiously. | .419 | |
| 239. My child is able to resist laughing or smiling when it isn't appropriate. | .458 | |
| 227. My child can easily stop an activity when s/he is told "no." | .715 | |
| 238. My child is good at games like "Simon Says," "Mother, May I?" and "Red Light, Green Light." | .480 | |
| 205. My child has difficulty waiting in line for something. | .641 | |
| 210. My child can wait before entering into new activities if s/he is asked to. | .766 | |
| 7. My child is good at following instructions. | .780 | |
| X01. My child prepares for trips and outings by planning things s/he will need. | .372 | |
| **6. My child has a hard time following instructions.** | **.766** | **.764** |
| **218. My child can lower his/her voice when asked to do so.** | **.661** | **.663** |
| **240. My child is usually able to resist temptation when told s/he is not supposed to do something.** | **.726** | **.666** |
| **223. My child has trouble sitting still when s/he is told to (at movies, church, etc.).** | **.649** | **.674** |
| R04. My child can easily stop a play activity when told to stop. | | .698 |
| 202. My child has an easy time waiting to open a present. | | .517 |
| 207. My child has a hard time waiting for his/her favorite snack or meal. | | .572 |
| 216. When I'm on a phone call and my child wants to talk to me, s/he will wait until I've finished the call. | | .559 |
| Model Fit | | |
| Chi-Square | 76.90 | 25.09 |
| df | 35 | 20 |
| p | < .0001 | .198 |
| RMSEA | .066 | .031 |
| 90% CI | .046 - .086 | .000 - .064 |
| CFI | .925 | .991 |

**Bolded** items appear on both versions.

Table 3.18

*Low Intensity Pleasure unconstrained CFA model.*

| | Standardized Loadings | |
|---|---|---|
| Items | Original | Revised |
| 247. My child enjoys "snuggling up" next to a parent. | .650 | |
| 270. My child likes being sung to. | .696 | |
| 277. My child doesn't enjoy being read to very much. | .487 | |
| 282. My child rarely enjoys just being talked to. | .219 | |
| 290. My child enjoys just sitting quietly in the sunshine. | .312 | |
| 293. My child enjoys looking at or reading books. | .517 | |
| 303. My child doesn't care much for quiet games. | .210 | |
| 308. My child isn't interested in watching quiet TV shows, such as "Mister Rogers." | .300 | |
| **255. My child enjoys sitting on a parent's lap.** | **.639** | **.638** |
| **259. My child enjoys gentle rhythmic activities, such as rocking or swaying.** | **.433** | **.487** |
| **283. My child enjoys just being talked to.** | **.496** | **.472** |
| **285. My child likes the sound of words, as in nursery rhymes or poems.** | **.487** | **.424** |
| **310. My child enjoys taking warm baths.** | **.389** | **.388** |
| 253. My child seeks opportunities to get hugs and kisses from family members. | | .450 |
| 273. My child enjoys listening to music quietly. | | .393 |
| 318. When around flowers, my child will take the time to smell them and enjoy their scent. | | .399 |
| R06. My child likes it when a parent or caregiver sings to him/her. | | .712 |
| Model Fit | | |
| Chi-Square | 172.71 | 49.93 |
| df | 65 | 27 |
| p | < .0001 | 0.005 |
| RMSEA | .078 | .056 |
| 90% CI | .064 - .092 | .031 - .080 |
| CFI | .780 | .917 |

**Bolded** items appear on both versions.

Table 3.19

*Perceptual Sensitivity unconstrained CFA model.*

| Items | Standardized Loadings | |
|---|---|---|
| | Original | Revised |
| 324. My child doesn't usually notice odors, such as perfume, smoke, cooking, etc. | .616 | |
| 338. My child usually comments if someone has an unusual voice. | .420 | |
| 354. My child notices the smoothness or roughness of objects s/he touches. | .575 | |
| 359. My child doesn't usually react to different textures of food. | .072[a] | |
| 382. My child doesn't seem to notice parents' facial expressions. | .334 | |
| 387. My child doesn't usually comment on people's facial features, such as size of nose or mouth. | .123[a] | |
| 363. My child is quickly aware of some new item in the living room. | .629 | |
| 369. My child notices even little specks of dirt on objects. | .387 | |
| 374. My child notices it when parents are wearing new clothing. | .655 | |
| 377. My child usually doesn't comment on changes in parents' appearance. | .638 | |
| **378. My child comments when a parent has changed his/her appearance.** | **.759** | **.483** |
| **331. My child seems to listen to even quiet sounds.** | **.481** | **.628** |
| 328. My child will notice different aromas in nature (e.g., flowers), even when the scents are subtle. | | .695 |
| 335. My child notices when very quiet music is playing in the background. | | .685 |
| 342. My child seems to notice when sounds change in volume, such as when someone turns the radio or television up or down. | | .640 |
| 353. My child seems to notice when a room is more warm or cool than expected. | | .529 |
| 358. My child likes to run his/her hand over things to see if they are smooth or rough. | | .402 |
| 380. My child notices small changes in the environment, like lights getting brighter in a room. | | .767 |
| Fit Indices | | |
| Chi-Square | 168.48 | 44.10 |
| df | 54 | 20 |
| p | < .0001 | .002 |
| RMSEA | .088 | .067 |

Table 3.19, continued

*Perceptual Sensitivity unconstrained CFA model (validation sample)*

| Fit Indices | Original | Revised |
|---|---|---|
| 90% CI | .073 - 0.104 | .040 - .093 |
| CFI | .801 | .947 |

[a] Non-significant, $p > .05$.

**Bolded** items appear on both versions.

lower standardized factor loadings ranging from $\lambda$ = .072 to $\lambda$ = .759, and two items with loadings less than the recommended threshold of .300 (Item 359 *“My child doesn't usually react to different textures of food.”*, $\lambda$ = .072; and Item 387 *“My child doesn't usually comment on people's facial features, such as size of nose or mouth.”*, $\lambda$ = .123). The original scale CFA demonstrated poor fit, $\chi^2(54) = 168.48$, $p < .0001$, RMSEA = .088 (90% CI: .073 - .104), CFI = .801.

***Scale Overlap.*** Additional evidence of content validity of the revised scales is provided by item overlap with the original CBQ scales: all revised scales contain at least two original CBQ EC scale items. The new Activation Control scale also contains two original CBQ items, one from the original Attentional Focusing scale (Item 30 / CBQ 16 *“When picking up toys or doing other jobs, my child usually keeps at the task until it's done.”*) and one from the original Attention Shifting scale (Item 99 / CBQ 180 *“My child has an easy time leaving play to come to dinner.”*).

The revised Attentional Focusing scale contains three original CBQ items (Item 130 / CBQ 38 *“When practicing an activity, my child has a hard time keeping her/his mind on it.”*; Item 143 / CBQ 171 *“My child is easily distracted when listening to a story.”*; and Item 162 / CBQ 186 *“My child sometimes becomes absorbed in a picture book and looks at it for a long time.”*). Furthermore, Item R01 (*“My child spends a long time engaged in drawing, coloring, or crafts.”*) is a rewrite of the original CBQ 144 (*“When building or putting something together, becomes very involved in what s/he is doing, and works for long periods.”*).

The revised Attention Shifting scale contains two original CBQ items (Item 79 / CBQ 184 *“My child sometimes doesn't seem to hear me when I talk to her/him.”*; and

Item 94 / CBQ 6 *"It is hard to get my child's attention when s/he is concentrating on something."*).

The revised Inhibitory Control scale contains four original CBQ items (Item 6 / CBQ 32 *"My child has a hard time following instructions."*; Item 218 / CBQ 4 *"My child can lower his/her voice when asked to do so."*; Item 223 / CBQ 108 *"My child has trouble sitting still when s/he is told to (at movies, church, etc.)"*; and Item 240 / CBQ 185 *"My child is usually able to resist temptation when told s/he is not supposed to do something."*). Furthermore, Item R04 (*"My child can easily stop a play activity when told to stop."*) is a rewrite of the original CBQ 168 (*"My child can easily stop an activity when s/he is told 'no.'"*).

The revised Low Intensity Pleasure scale contains five original CBQ items (Item 255 / CBQ 174 *"My child enjoys sitting on a parent's lap."*; Item 259 / CBQ 164 *"My child enjoys gentle rhythmic activities, such as rocking or swaying."*; Item 283 / CBQ 113 *"My child enjoys just being talked to."*; Item 285 / CBQ 151 *"My child likes the sound of words, as in nursery rhymes or poems."*; and Item 310 / CBQ 54 *"My child enjoys taking warm baths."*). Furthermore, Item R06 (*"My child likes it when a parent or caregiver sings to him/her."*) is a rewrite of CBQ 146 (*"My child likes being sung to."*).

The revised Perceptual Sensitivity scale contains two original CBQ items (Item 331 / CBQ 52 *"My child seems to listen to even quiet sounds."*; Item 378 / CBQ 65 *"My child comments when a parent has changed his/her appearance."*).

***Dimensionality.*** Dimensionality was assessed using R version 3.4.0 (R Core Team, 2017) and the package "nFactors" (Raiche, 2010). Figure 3.2 through Figure 3.7 are Scree plots for revised scales provided by nFactors, and includes the number of

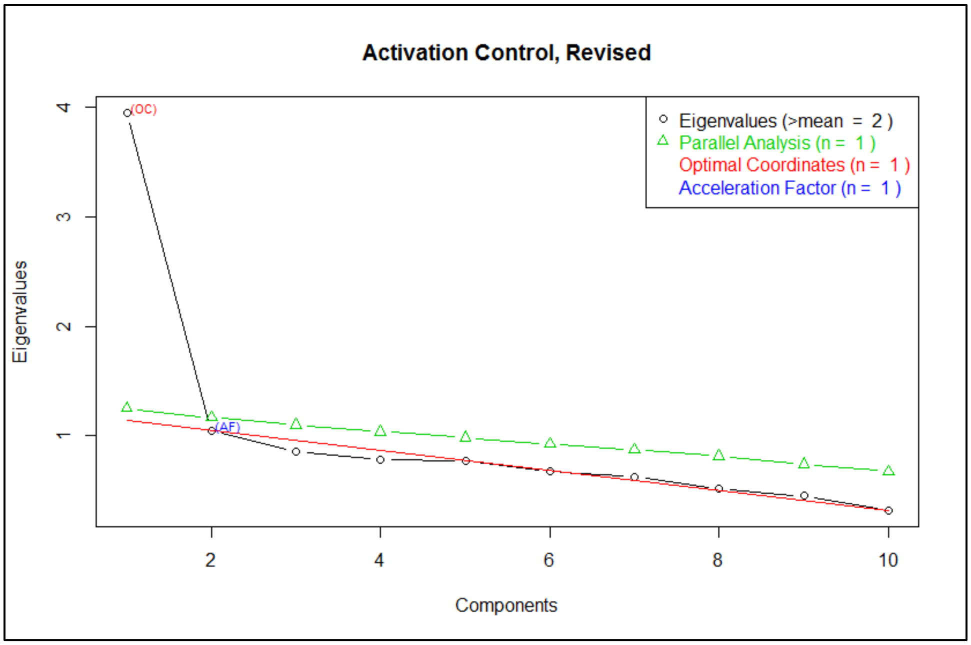



Figure 3.2
*Dimensionality of CBQ-R Activation Control scale*

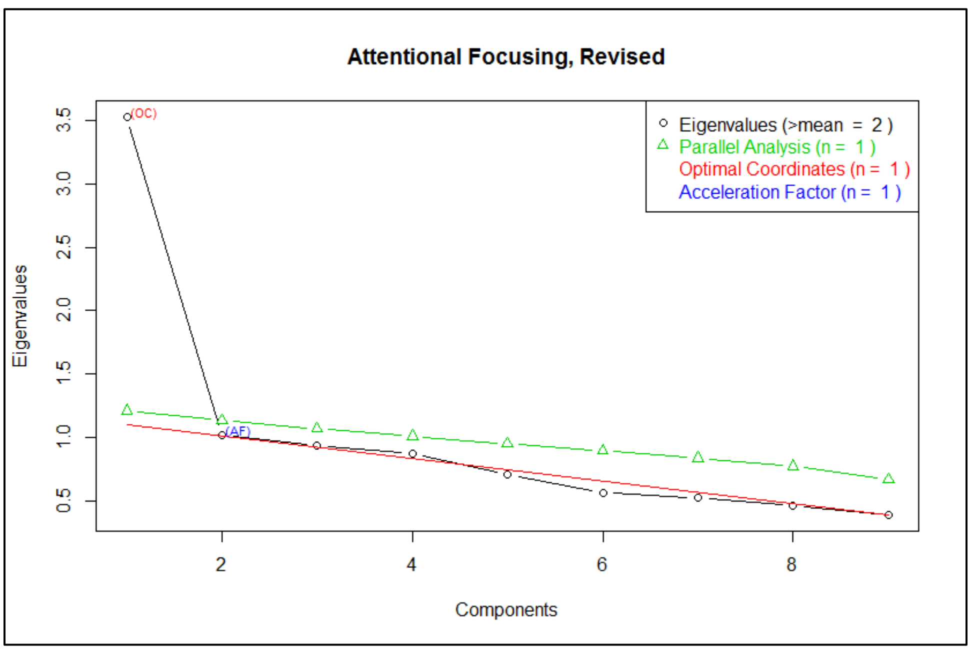



Figure 3.3
*Dimensionality of CBQ-R Attentional Focusing Scale*

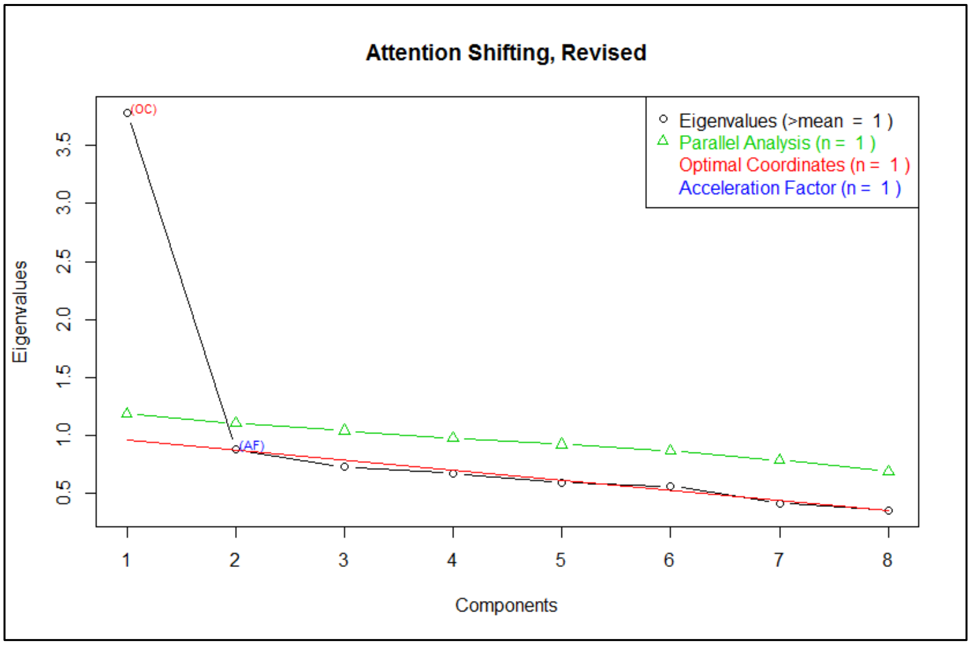



Figure 3.4
*Dimensionality of CBQ-R Attention Shifting Scale*

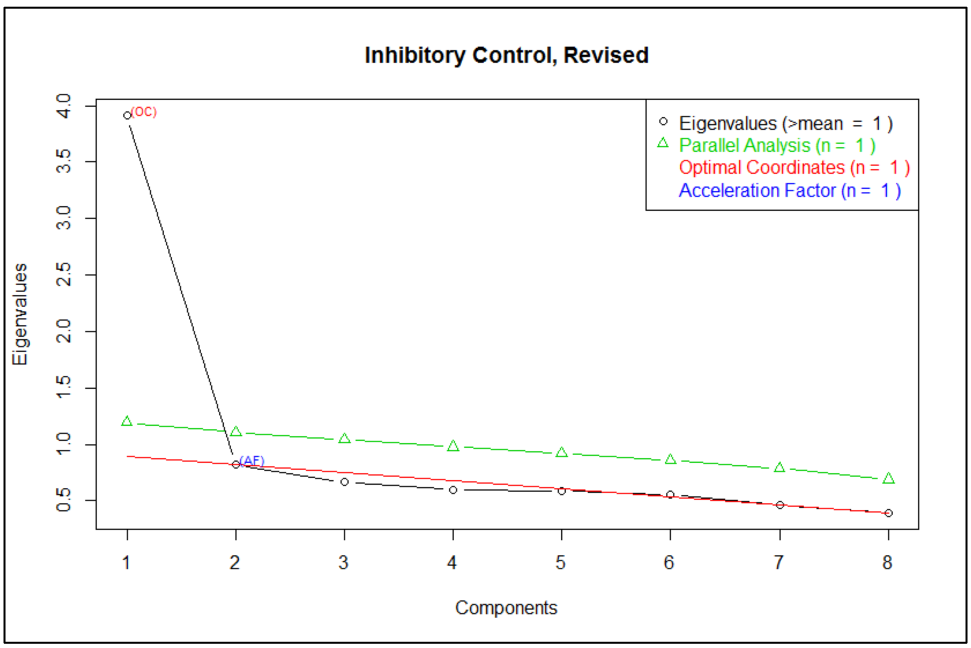



Figure 3.5
*Dimensionality of CBQ-R Inhibitory Control Scale*

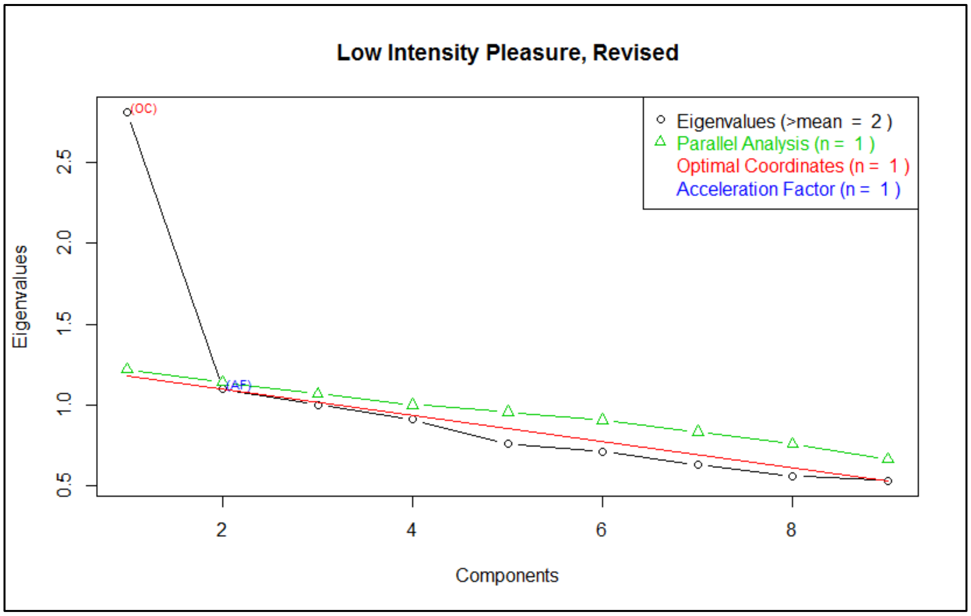



Figure 3.6
*Dimensionality of CBQ-R Low Intensity Pleasure Scale*

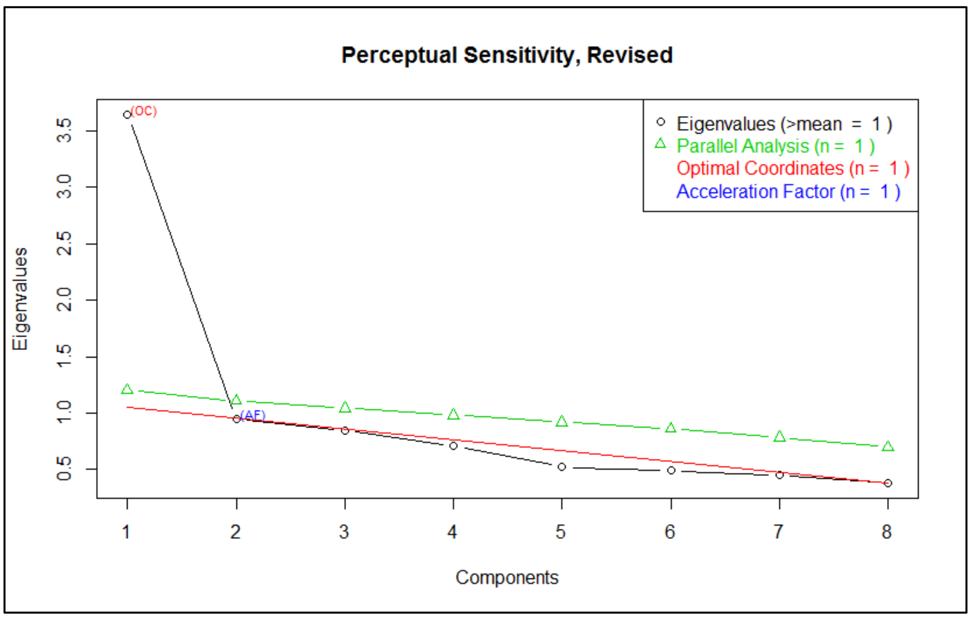



Figure 3.7
*Dimensionality of CBQ-R Perceptual Sensitivity Scale*

estimated factors based on four tests of dimensionality: Kaiser-Guttman Rule (labeled as "Eigenvalues"), Parallel Analysis, Optimal Coordinates, and Acceleration Factor). All four tests recommended a single factor for three revised scales: Attention Shifting (Figure

3.4), Inhibitory Control (Figure 3.5), and Perceptual Sensitivity (Figure 3.7). For the other revised scales (Activation Control, Figure 3.2; Attentional Focusing, Figure 3.3; Low Intensity Pleasure, Figure 3.6), all tests recommended a single factor except for the Kaiser-Guttman test, which recommended two factors. This provides sufficient evidence of unidimensionality, given most tests indicated each scale represents a single factor.

In contrast, Scree plots for original CBQ scales are presented in Figure 3.8 through Figure 3.12. The original Attentional Focusing scale (Figure 3.8) appears to be comprised of two factors, based on the Kaiser-Guttman rule, Parallel Analysis, and Optimal Coordinates (although the Acceleration Factor still indicates a single factor). The original Attention Shifting scale (Figure 3.9) seems to be unidimensional. The original Inhibitory Control scale (Figure 3.10) also seems to be unidimensional, with only the Kaiser-Guttman rule suggesting multiple factors. The Original Low Intensity Pleasure scale (Figure 3.11) may be multidimensional, with two tests indicating multiple factors: the Kaiser Guttman rule (four factors), and Parallel Analysis (three factors). Perceptual Sensitivity (Figure 3.12) also appears to be multidimensional, with only the Acceleration Factor suggesting unidimensionality, while Parallel Analysis and Optimal Coordinates indicate two factors, and the Kaiser-Guttman rule indicate three.

**Additional analyses**. Response frequency analysis using the validation sample is presented in Table 3.20, with percentages indicating the proportion of items on the scale for which each response option (1 through 7) had low response use (< 5%). Overall, the original and revised versions have similar patterns of low response use. The lowest response category—corresponding to “Never”—was infrequently used, with nearly all items having low response use for each scale. The highest response category—

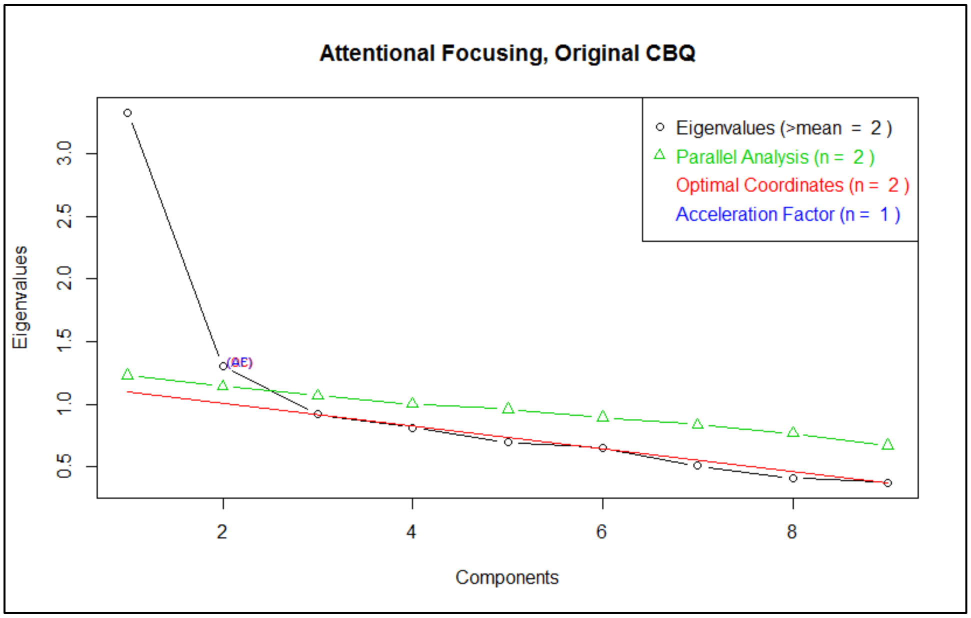



Figure 3.8
*Dimensionality of Original CBQ Attentional Focusing Scale*

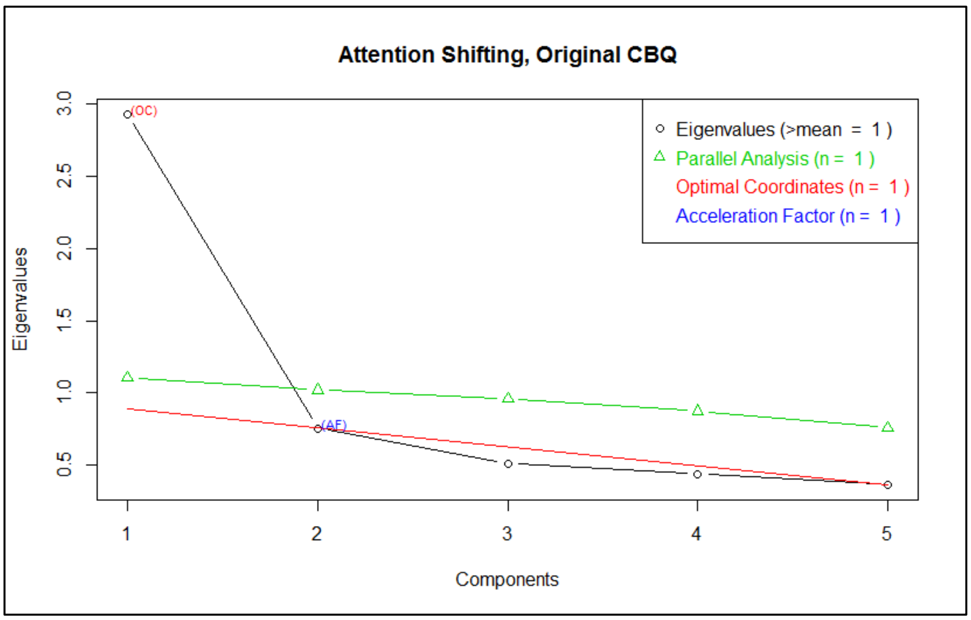



Figure 3.9
*Dimensionality of Original CBQ Attention Shifting Scale*

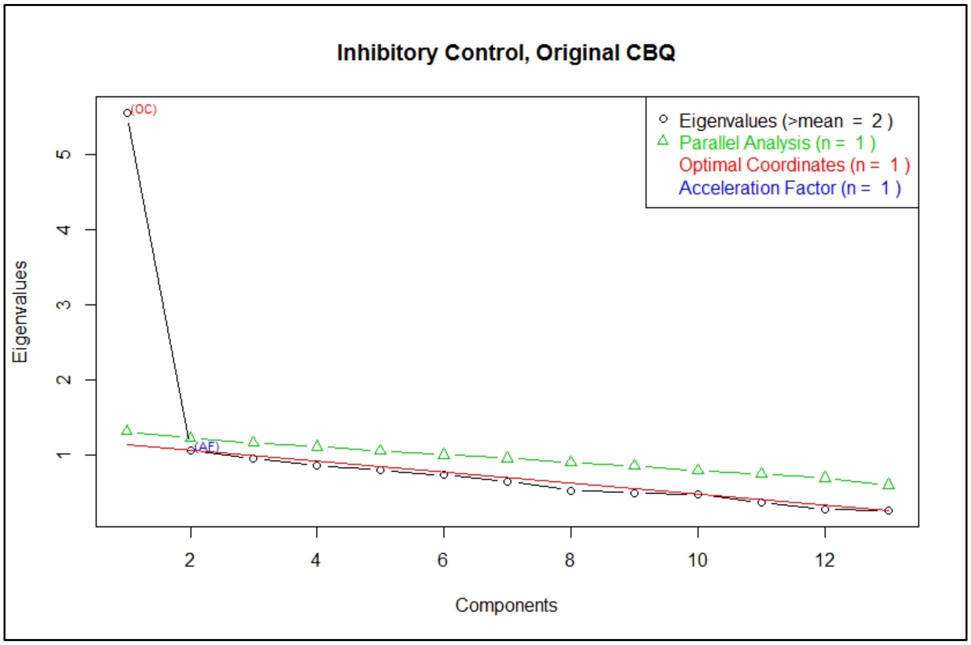



Figure 3.10

*Dimensionality of Original CBQ Inhibitory Control Scale*

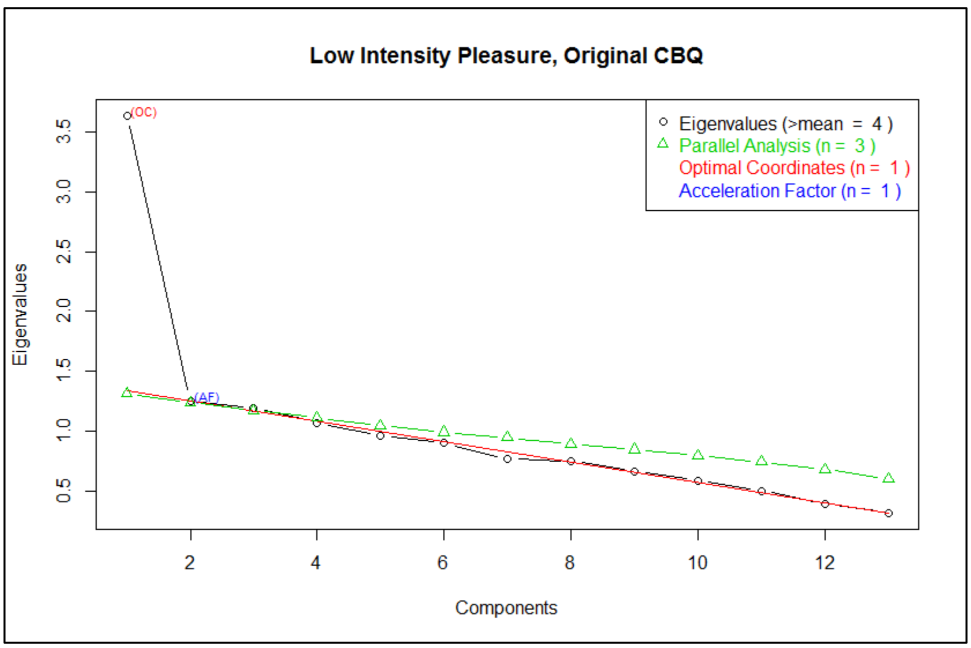



Figure 3.11

*Dimensionality of Original CBQ Low Intensity Pleasure Scale*

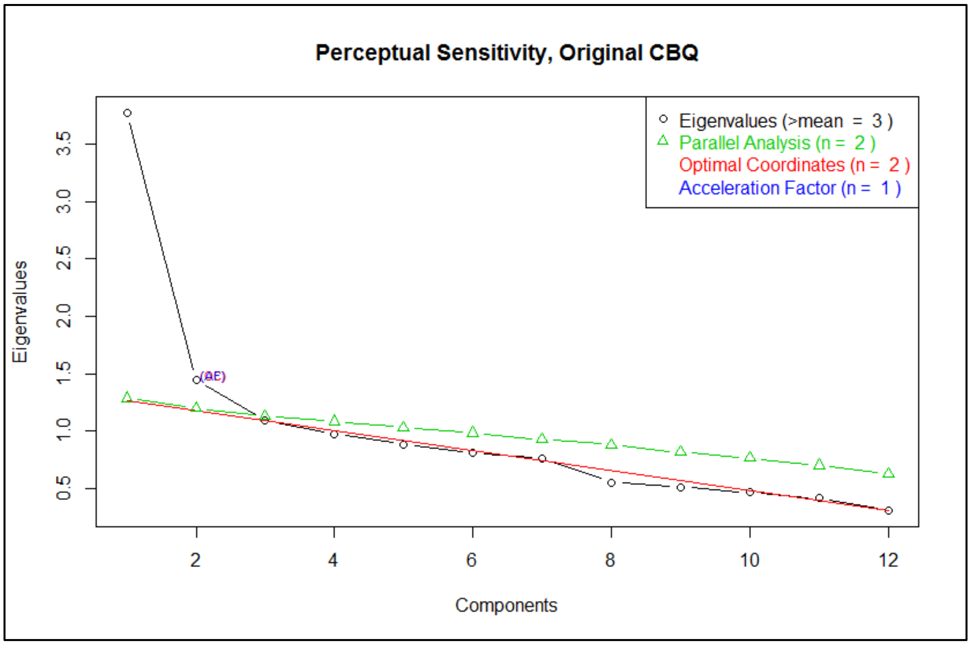



Figure 3.12

*Dimensionality of Original CBQ Perceptual Sensitivity Scale*

Table 3.20

*Percent original and revised CBQ Effortful Control scales with low response option use (<5%).*

| | Response Option* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Activation Control - Revised | 100.0% | 10.0% | 0.0% | 0.0% | 0.0% | 0.0% | 80.0% |
| Attentional Focusing - Original | 100.0% | 44.4% | 11.1% | 0.0% | 0.0% | 0.0% | 55.6% |
| Attentional Focusing - Revised | 100.0% | 55.6% | 11.1% | 0.0% | 0.0% | 0.0% | 33.3% |
| Attention Shifting - Original | 100.0% | 40.0% | 0.0% | 0.0% | 0.0% | 0.0% | 100.0% |
| Attention Shifting - Revised | 87.5% | 37.5% | 0.0% | 0.0% | 0.0% | 0.0% | 87.5% |
| Inhibitory Control - Original | 92.3% | 38.5% | 7.7% | 0.0% | 0.0% | 0.0% | 46.2% |
| Inhibitory Control - Revised | 87.5% | 12.5% | 0.0% | 0.0% | 0.0% | 0.0% | 87.5% |
| Low Intensity Pleasure - Original | 92.3% | 76.9% | 61.5% | 38.5% | 0.0% | 0.0% | 15.4% |
| Low Intensity Pleasure - Revised | 100.0% | 44.4% | 55.6% | 22.2% | 0.0% | 0.0% | 0.0% |
| Perceptual Sensitivity - Original | 83.3% | 16.7% | 16.7% | 8.3% | 0.0% | 0.0% | 16.7% |
| Perceptual Sensitivity - Revised | 100.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 12.5% |
| All Items - Original | 96.2% | 26.9% | 11.5% | 3.8% | 0.0% | 0.0% | 50.0% |
| All Items - Revised | 92.3% | 44.2% | 23.1% | 11.5% | 0.0% | 0.0% | 38.5% |

* Accounting for reverse-scored items

| Color scale: | 100% | > 75% | > 50% | > 25% |
|---|---|---|---|---|

corresponding to “Always”—had considerable low usage in the revised Activation Control, Attention Shifting, and Inhibitory Control scales, as well as the original Attention Shifting scale. The revised Low Intensity Pleasure scale seemed to be a slight improvement, with response options 2 through 4—“Very Rarely”, “Less than Half the Time”, and “Half the Time”—more frequently used than the for the original scale. In general, however, the spread of responses for all revised scales were comparable to their original counterparts.

Independent samples t-tests using the calibration sample are presented in Table 3.21. I used a Bonferroni correction to control for familywise Type I error by lowering the threshold by which a given t-value would be considered significant (13 hypothesis tests; $\alpha = .0038$). For the revised scales, older children were on average rated significantly higher than younger children on Activation Control scale, $M_{young} = 4.30$ vs. $M_{old} = 4.73$, $t = 5.56$, $p < .001$, Cohen’s $d = .556$, Attentional Focusing, $M_{young} = 4.83$ vs. $M_{old} = 5.11$, $t = 3.80$, $p < .001$, Cohen’s $d = .380$, Inhibitory Control, $M_{young} = 4.26$ vs. $M_{old} = 4.72$, $t = 5.17$, $p < .001$, Cohen’s $d = .517$, and EC overall, $M_{young} = 4.69$ vs. $M_{old} = 4.90$, $t = 3.61$, $p < .001$, Cohen’s $d = .361$. For the original scales, older children were rated as significantly higher than younger children on Attentional Focusing, $M_{young} = 4.67$ vs. $M_{old} = 4.94$, $t = 3.59$, $p < .001$, Cohen’s $d = .359$, Attention Shifting, $M_{young} = 4.42$ vs. $M_{old} = 4.65$, $t = 2.48$, $p = .014$, Cohen’s $d = .248$, Inhibitory Control, $M_{young} = 4.70$ vs. $M_{old} = 5.16$, $t = 5.84$, $p < .001$, Cohen’s $d = .584$, and EC overall, $M_{young} = 4.91$ vs. $M_{old} = 5.07$, $t = 3.15$, $p = .002$, Cohen’s $d = .315$. Younger children were rated higher on the original Low Intensity Pleasure scale than older children, $M_{young} = 5.61$ vs. $M_{old} = 5.47$, $t = -2.29$, $p = .022$, Cohen’s $d = .229$. There were no significant ($p < .0038$) differences in group means for the revised Attention Shifting scale, $M_{young}$ = 4.52 vs. $M_{old}$ = 4.71, $t$ = 2.16, $p$ = .031, Cohen's $d$ = .216, original Attention Shifting scale, $M_{young}$ = 4.42 vs. $M_{old}$ = 4.65, $t$ = 2.48, $p$ = .014, Cohen's $d$ = .248, revised Low Intensity Pleasure scale, $M_{young}$ = 5.53 vs. $M_{old}$ = 5.43, $t$ = -1.38, $p$ = .168, Cohen's $d$ = .138, original Low Intensity Pleasure scale, $M_{young}$ = 5.61 vs. $M_{old}$ = 5.47, $t$ = -2.29, $p$ = .022, Cohen's $d$ = .229, revised Perceptual Sensitivity scale, $M_{young}$ = 4.65 vs. $M_{old}$ = 4.66, $t$ = 0.06, $p$ = .953, Cohen's $d$ = .006, or original Perceptual Sensitivity scale, $M_{young}$ = 4.75 vs. $M_{old}$ = 4.82, $t$ = 0.96, $p$ = .335, Cohen's $d$ = .096.

Table 3.21

*Independent samples t-tests for original and revised CBQ Effortful Control scales with effect sizes.*

| Scale | Young Mean (*SD*) | Old Mean (*SD*) | *t* [a] | *p* | *d* |
|---|---|---|---|---|---|
| Activation Control - Revised | 4.30 (0.78) | 4.73 (0.79) | 5.56 | < .001 | .556 |
| Attentional Focusing - Original | 4.67 (0.72) | 4.94 (0.75) | 3.59 | < .001 | .359 |
| Attentional Focusing - Revised | 4.83 (0.72) | 5.11 (0.78) | 3.80 | < .001 | .380 |
| Attention Shifting - Original | 4.42 (0.87) | 4.65 (0.96) | 2.48 | .014 | .248 |
| Attention Shifting - Revised | 4.52 (0.86) | 4.71 (0.88) | 2.16 | .031 | .216 |
| Inhibitory Control - Original | 4.70 (0.79) | 5.16 (0.79) | 5.84 | < .001 | .584 |
| Inhibitory Control - Revised | 4.26 (0.91) | 4.72 (0.89) | 5.17 | < .001 | .517 |
| Low Intensity Pleasure - Original | 5.61 (0.55) | 5.47 (0.63) | -2.29 | .022 | .229 |
| Low Intensity Pleasure - Revised | 5.53 (0.69) | 5.43 (0.76) | -1.38 | .168 | .138 |
| Perceptual Sensitivity - Original | 4.75 (0.78) | 4.82 (0.75) | 0.96 | .335 | .096 |
| Perceptual Sensitivity - Revised | 4.65 (0.95) | 4.66 (0.99) | 0.06 | .953 | .006 |
| All Items - Original | 4.91 (0.52) | 5.07 (0.54) | 3.15 | .002 | .315 |
| All Items - Revised | 4.69 (0.59) | 4.90 (0.62) | 3.61 | < .001 | .361 |

[a] *df* ranged from 390.67 to 397.97.

Note: Calibration sample

# CHAPTER 4

# DISCUSSION

The revised CBQ EC scales generally show marked improvement over the original versions. In this section, I will review the technical and theoretical comparisons of the original and revised scales, discuss implications of revising a measure of effortful control to the study of child temperament and development, present limitations to the utility of the revised scales, and suggest future directions for improving the way in which we measure effortful control.

## The Revised Scales

At the end of Chapter 2, I outlined three criteria the newly revised EC scales must meet to be considered an improvement upon the original versions. First, the revised scales must accurately represent the constructs of effortful control. Second the revised scales should show measurement invariance across ages 3-7. Third, the scales should demonstrate strong psychometric properties under multiple measurement frameworks (i.e., CTT, CFA). In this section I discuss each of these criteria in relation to the original and revised CBQ EC scales.

### Construct Representation

Rothbart and colleagues have outlined components central to effortful control as the abilities to (a) suppress or inhibit a dominant response, (b) perform a subdominant response, (c) plan actions, (d) detect errors, (e) focus attention, and (f) shift attention (Rothbart et al., 2001; Rothbart & Bates, 2006; Rueda, 2012). Each revised scale contains content related primarily to the construct to be measured, and aligns with the finalized scale definitions. This was affirmed by content experts upon reviewing and modifying the statistically refined scales and accepting the final versions. The content experts and I collectively decided against including *planning actions* in the revised EC scales, as it merely requires EC, rather than serving as a core component. The remaining five components are directly measured by five of the six scales. Inhibitory Control explicitly measures suppressing or inhibiting a dominant response. Activation Control measures performing a subdominant action. Attentional Focusing obviously measures focused attention, as does Attention Shifting measure shifting attention. Perceptual Sensitivity measures sensitivity to subtle stimuli, including the ability to detect small changes in the environment which relates to the detection of errors. The remaining Low Intensity Pleasure scale does not seem to directly relate to any of Rothbart's proposed components of EC—although the original scale did not either. The construct of Low Intensity Pleasure does not appear to be necessarily *voluntary*, a defining aspect of EC (Rueda, 2012). For instance, a child high in Low Intensity Pleasure may respond with pleasure to the feel of a warm bath—as measured via Item 310 / CBQ 54—but that pleasure is not a voluntary control of emotion, but rather a reaction to low intensity stimuli. Further, it could be argued that Low Intensity Pleasure perhaps relates to the factor of EC as a secondary component; it is not a core feature of EC, but instead EC somewhat facilitates the capacity to derive pleasure from low intensity stimuli. Additional considerations are discussed in the Implications / Disadvantages section of this chapter.

Each revised scale is also clearly unidimensional. Of the four employed methods for ascertaining dimensionality, all of the revised scales are unidimensional based on three or four methods. The same cannot be said about the original scales, with two or more methods suggesting the original Attentional Focusing, Low Intensity Pleasure, and Perceptual Sensitivity scales are multidimensional (for overview, see: Table 4.1). Multidimensionality indicates that the relationship among items is not explained by a single factor and thus, the scale measures more than one construct. Furthermore, multidimensionality negatively impacts estimates of reliability (e.g., $\alpha$; Cortina, 1993) and is therefore an undesirable property for a scale typically used as a unidimensional measure of a construct, as are the scales of the CBQ.

Table 4.1

*Overview of dimensionality of original and revised CBQ Effortful Control scales*

| | Dimensionality (N Factors) | | | |
|---|---|---|---|---|
| Scale | KG | PA | OC | AF |
| Activation Control - Revised | 2 | 1 | 1 | 1 |
| Attentional Focusing - Original | 2 | 2 | 2 | 1 |
| Attentional Focusing - Revised | 2 | 1 | 1 | 1 |
| Attention Shifting - Original | 1 | 1 | 1 | 1 |
| Attention Shifting - Revised | 1 | 1 | 1 | 1 |
| Inhibitory Control - Original | 2 | 1 | 1 | 1 |
| Inhibitory Control - Revised | 1 | 1 | 1 | 1 |
| Low Intensity Pleasure - Original | 4 | 3 | 1 | 1 |
| Low Intensity Pleasure - Revised | 2 | 1 | 1 | 1 |
| Perceptual Sensitivity - Original | 3 | 2 | 2 | 1 |
| Perceptual Sensitivity - Revised | 1 | 1 | 1 | 1 |

Note: Data are Study 2 validation sample

KG = Kaiser Guttman Rule, PA = Parallel Analysis, OC = Optimal Coordinates, AF = Acceleration Factor

**Measurement Invariance**

An overview of comparisons of measurement invariance using Study 2 calibration sample are presented in Table 4.2. All Activation Control items were metric invariant, and six of the 10 items were scalar invariant. All revised Attentional Focusing items were metric invariant, and seven of the nine items were scalar invariant. For the original Attentional Focusing scale, results indicated one of nine items was configurally noninvariant and three items were scalar noninvariant. All revised Attention Shifting items were metric invariant, and seven of eight items were scalar invariant. Of the original five Attention Shifting items, two displayed scalar noninvariance. All revised Inhibitory Control items were metric and scalar invariant. For the original 13-item Inhibitory Control scale, two items were scalar noninvariant. All revised Low Intensity Pleasure items were metric and scalar invariant. Of the original 13-item Low Intensity Pleasure scale, five were flagged for configural noninvariance. All revised Perceptual Sensitivity items were metric invariant, and seven of nine items were scalar invariant. Of the original 12 Perceptual Sensitivity items, two were flagged for configural noninvariance, and four were flagged for scalar noninvariance.

The revised scales thus represent an improvement over the original scales in terms of age-group measurement invariance, as the original scales show comparatively more noninvariance than the revised scales. Furthermore, the revised scales show only some indication of noninvariance in the form of scalar noninvariance, meaning that the structure of the construct and how items relate to it are consistent for younger and older children (configural and metric invariance, respectively). While four of the new scales show some degree of scalar noninvariance, the proportion of scalar invariant items is

Table 4.2

*Overview of CFA noninvariance of original and revised CBQ EC scales.*

| | Measurement Noninvariance (% items) | | | |
|---|---|---|---|---|
| Scale | Configural | Metric | Scalar | **Overall** |
| Activation Control - Revised | 0.0% | 0.0% | 40.0% | **40.0%** |
| Attentional Focusing - Original | 11.1% | 0.0% | 33.3% | **44.4%** |
| Attentional Focusing - Revised | 0.0% | 0.0% | 22.2% | **22.2%** |
| Attention Shifting - Original | 0.0% | 0.0% | 40.0% | **40.0%** |
| Attention Shifting - Revised | 0.0% | 0.0% | 12.5% | **12.5%** |
| Inhibitory Control - Original | 0.0% | 0.0% | 15.4% | **15.4%** |
| Inhibitory Control - Revised | 0.0% | 0.0% | 0.0% | **0.0%** |
| Low Intensity Pleasure - Original | 38.5% | 0.0% | 0.0% | **38.5%** |
| Low Intensity Pleasure - Revised | 0.0% | 0.0% | 0.0% | **0.0%** |
| Perceptual Sensitivity - Original | 16.7% | 0.0% | 33.3% | **50.0%** |
| Perceptual Sensitivity - Revised | 0.0% | 0.0% | 22.2% | **22.2%** |

Note: Data are Study 2 calibration sample.

higher than for the original version. This indicates that, although mean comparisons by age-group may be affected by measurement error, this effect may be less prominent with the revised scales.

**Psychometric Properties**

An overview of the Classical Test Theory internal consistency and Confirmatory Factor Analysis fit indices of original and revised scales with coefficient interpretation (e.g., Cronbach's $\alpha > .80$ is "Very Good") is presented in Table 4.3. Based on conventional interpretation of $\alpha$, omega, and unstructured CFA fit indices, the revised scales performed as well, if not better than, the original scales for every coefficient examined.

Table 4.3

*Overview of CTT internal consistency and CFA fit indices for original and revised CBQ-R EC scales.*

| | CTT[a] | | CFA[b] | | |
|---|---|---|---|---|---|
| Scale | $\alpha$ | Omega | $\chi^2$ | RMSEA | CFI |
| Activation Control - Revised | Very Good | Very Good | Good | Acceptable | Acceptable |
| Attentional Focusing - Original | Respectable | Respectable | Not Good | Mediocre | Unacceptable |
| Attentional Focusing - Revised | Respectable | **Very Good** | Not Good | **Acceptable** | **Acceptable** |
| Attention Shifting - Original | Very Good | Very Good | Good | Mediocre | Good |
| Attention Shifting - Revised | Very Good | Very Good | Good | **Close** | Good |
| Inhibitory Control - Original | Very Good | Very Good | Not Good | Acceptable | Acceptable |
| Inhibitory Control - Revised | Very Good | Very Good | **Good** | **Close** | **Good** |
| Low Intensity Pleasure - Original | Respectable | Respectable | Not Good | Mediocre | Unacceptable |
| Low Intensity Pleasure - Revised | Respectable | Respectable | Not Good | **Close** | **Good** |
| Perceptual Sensitivity - Original | Respectable | Respectable | Not Good | Mediocre | Unacceptable |
| Perceptual Sensitivity - Revised | **Very Good** | **Very Good** | Not Good | **Acceptable** | **Acceptable** |
| Interpretation | | | | | |
| (From worse to better) | Poor: < .60<br>Mediocre: .60 - .70<br>Respectable: .70 - .80<br>Very Good: > .80 | | Not Good: $p < .01$<br>Good: $p > .01$ | Unacceptable: > .10<br>Mediocre: .08 - .10<br>Acceptable: .06 - .08<br>Close: < .06 | Unacceptable: < .90<br>Acceptable: .90 - .95<br>Good: > .95 |

[a] Source: DeVellis (2017)

[b] Source: Brown (2006), Hu & Bentler (1998)

Note: Validation Sample. **Bolded** values indicate when fit interpretation for the revised scale was better than the original scale. No fit interpretations for the original were better than the revised.

For four of the original five scales, the α and omega values of the original and revised versions were similar (at least in interpretation). The revised version of Perceptual Sensitivity had very good CTT internal consistency, whereas the original version only had acceptable reliability. The Revised Attentional Focusing scale had very good internal consistency compared to the original according to omega, but α values for both versions were in the acceptable range. Furthermore, when examining Table 4.3, we see that for all but one scale, the mean inter-item correlation for the revised version is higher than the original version. The exception here is the Attention Shifting scale, as the revised version contains three more items than the original, yet the internal consistency remained about the same. Examining the overall magnitude of the internal consistency values in Table 3.11, we see that all revised scales had higher α values than the originals with the exception of the Low Intensity Pleasure scale, although the revised Low Intensity Pleasure scale has four fewer items than he original.

The three CFA fit indices each give a separate indication of how well the model fits the data. The $\chi^2$ statistic represents the difference between the model-implied covariance matrix (i.e., what the matrix would look like based solely on our parameter estimates) and the observed covariance matrix (Brown, 2006). A model-implied covariance matrix that is significantly different from the observed covariance matrix (ie., it doesn't recreate the actual covariance matrix very well), indicates poor fit. However, as mentioned before, the $\chi^2$ statistic is very sensitive to sample size, and with larger samples (e.g., $N > 200$) as in this case, the statistic is typically significant at $\alpha = .05$. As you may notice, only four scales have non-significant $\chi^2$ statistics, three of which are the revised versions. Given the large sample size, this shows that the CFA for the revised scales reproduce the covariance matrices well without additional correlated error terms, suggesting there is little overlap among items not accounted for by the latent factor.

Root Mean Square Error of Approximation (RMSEA) also indicates how well a model reproduces the observed covariance matrix, but penalizes for model complexity. For example, whereas a CFA with a lot of correlated error terms may fit well according to $\chi^2$—because a lot of correlated errors help better reproduce the covariance matrix—the RMSEA may still indicate that the model fits poorly, especially if some of the correlated errors are superfluous. Thus, RMSEA is an index of *parsimony*, indicating whether the model fits the population *reasonably* well, whereas $\chi^2$ as an index of *absolute fit* and indicates if the model fits the population perfectly. For the three revised scales where the $\chi^2$ statistic was non-significant (Attentional Focusing, Low Intensity Pleasure, and Perceptual Sensitivity), the RMSEA value was either close or acceptable. This suggests that although the $\chi^2$ statistic was non-significant—perhaps due to sample size—and thus the current model does not fit perfectly to our data, it fits the data *reasonably* well. In comparison, four of the original scales demonstrate mediocre fit according to RMSEA. Because the CFA models summarized in Table 4.3 are estimated with no correlated errors, it is likely that observing a significant $\chi^2$ statistic and mediocre RMSEA is because the models do not adequately reproduce the covariance matrices, rather than being too complex. In comparison, the Attention Shifting scale has a non-significant $\chi^2$ statistic (at $\alpha = .01$), but only a mediocre RMSEA, suggesting the model does not reproduce the covariance matrix well given the available degrees of freedom. The original Attention Shifting scale also has only five items, which may leave too few remaining degrees of freedom for RMSEA to be adequately estimated and interpreted (see: Kenny, Kaniskan, & McCoach, 2014).

Unlike $\chi^2$ and RMSEA, which indicate how closely a model fits to observed data, the Comparative Fit Index (CFI) gives an indication of how much better a model fits the data compared to a covariance matrix with all covariance terms fixed to zero (Brown, 2006). In other words, $\chi^2$ and RMSEA tell us *how close our model is to the best model*, while CFI tells us *how much better our model is compared to a terrible model*. CFI also penalizes for model complexity. Of the revised scales, three demonstrated acceptable CFI fit, and three good CFI fit. Of the original scales, however, three showed unacceptable CFI fit. These three scales (Attentional Focusing, Low Intensity Pleasure, and Perceptual Sensitivity) all showed poor to mediocre psychometric properties in both CTT and CFA frameworks, whereas the revised versions showed considerably better properties.

Response frequency analyses showed comparable patterns of low response option use between original and revised EC scales, although there is slight indication the revised scale provides more diffuse response option selection than the original scale. As indicated in Study 1, the original Low Intensity Pleasure in particular suffered from a narrow range of response options being selected (mainly options 5 through 7). It appears this issue is somewhat ameliorated in the revised version.

Independent samples t-tests show that for most original and revised scales, older children are rated significantly higher than younger children. However, as found in Study 1, ratings on the original Low Intensity Pleasure scale in the present study are on average lower for older children than younger children. This contradicts the general understanding of the developmental nature of effortful control: EC is the control of behavior, attention, and affect, and this voluntary capacity is found to increase over the course of childhood (Carlson, 2005; Kochanska, 2000; Rueda, 2012). Scores on the revised Low Intensity Pleasure scale did not differ between age groups, however, providing some theoretical validity to the new version. Group differences were not found for either version of the Perceptual Sensitivity scale.

Overall, the revised scales CBQ EC scales show much improvement over the original versions. The revised scales more accurately represent their defined constructs and the parent factor of effortful control than the original version. The revised scales are all likely unidimensional, whereas some original scales seem to be multidimensional. The revised scales show a much greater degree of measurement invariance than the original scales. The revised scales also have generally better psychometric properties than their original counterparts, in both CTT and CFA contexts.

**Implications of Revised CBQ Effortful Control Scales**

In this section, I briefly overview the implications of the revised scales in terms of advantages and disadvantages, and discuss them in greater detail below. The revised CBQ Effortful Control scales were constructed with the intent to improve upon the reliability and validity of the original, with particular attention to the issue of longitudinal measurement invariance. Given the psychometric evaluation of both versions, the revised EC scales appear to be generally superior to the originals and therefore present a number of advantages to researchers and the study of temperament. The shorter length of each scale, with the exceptions of Attention Shifting and Attentional Focusing, will reduce the participant burden for researchers interested in using only certain scales of the revised CBQ, a common practice for researchers studying EC (e.g., Diaz et al., 2016; Eisenberg et al., 2007; Valiente et al., 2003). The new scales will also aid our understanding of the nature and development of effortful control throughout childhood by more accurately measuring constructs through the reduction of measurement error, allowing for better longitudinal investigations of temperament, and providing a broader picture of effortful control through the addition of the Activation Control scale. Furthermore, revised items were edited with explicit considerations of age, geography, culture, and native language, and therefore it is possible the revised CBQ EC items will be interpreted more consistently across a variety of contexts. Finally, while he revised scales will not necessarily discredit prior research on EC, they will hopefully reduce the prevalence of discrepant findings in the future.

While there are numerous advantages to a better measure of EC, there are a few shortcomings of revised scales. First, the revised Low Intensity Pleasure scale is still a bit problematic, and initial concerns (see Chapter 3, Step 2: Define Constructs) about the construct remain. Second, the CBQ is fairly ubiquitous in temperament and self-regulation literature, with the same instrument used for about the last 20 years and in dozens of countries. The introduction of new scales requires acceptance and adoption within the research community, and may impact our understanding of the nature of effortful control. Third, the revised scales still rely on parent report of children's temperament, and therefore will suffer from some bias associated with this source.

**Advantages of the Revised CBQ Effortful Control Scales**

On the surface, the revised scales have two desirable properties for researchers compared to the original scales: shorter length and better psychometric properties. While the overall revised EC scales contain as many items as the original (52 items), the revised version contains an additional scale (Activation Control) and shorter Inhibitory Control, Low Intensity Pleasure, and Perceptual Sensitivity scales. Shorter scales are desirable because they place less burden on respondents (DeVellis, 2017) and may allow researchers to free up time for the administration of other questionnaires, especially in large-scale studies. For example, in the longitudinal National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD), the study investigators chose to only use two shortened CBQ effortful control scales for parent report at age 54 months: Inhibitory Control (10 of 13 items) and Attentional Focusing (8 of 9 items; "Phase II Instrument Document", n.d.). It is likely the decision to use such a modified measure of effortful control was due to space constraints, as parents participated in or responded to over 20 interviews and questionnaires at this time point.

Putnam and Rothbart (2006) recognized the need to shorten the full version, producing short and very short forms of the instrument. However, these versions are not without shortcomings. Compared to the original version, the four EC scales on the CBQ Short Form (Attentional Focusing, Inhibitory Control, Low Intensity Pleasure, and Perceptual Sensitivity) represent a 45% reduction in scale length, yet only one short form CBQ scale improved alpha (Attentional Focusing, .73 to .75) and the other three had diminished alphas by an average of .06 (Putnam & Rothbart, 2006). The Very Short Form (CBQ-VSF) EC scale represents a 74% reduction in scale length compared to the original CBQ. Although its internal consistency is reported as acceptable ($\alpha \approx .78$; Putnam & Rothbart, 2006), evaluations of the VSF reveal problems with its factor structure, with EC items loading on non-EC factors, and non-EC items loading on the EC factor (Allan et al., 2013; de la Osa et al., 2014). This highlights the tradeoff faced in scale development: decreasing the length can negatively impact reliability and validity. The present study faced the same issues, although the advantage of revising scales with a large item pool is that I was able to replace poorer functioning items with better ones. Putnam and Rothbart (2006) did not take this approach, and thus were restricted into working with only items that appeared in the original version.

The revised scales also demonstrate favorable psychometric properties. Two scales have alphas between .70 and .80 (Attentional Focusing and Low Intensity Pleasure), and the four other scales have alphas above .80, whereas only one revised scale, Low Intensity Pleasure, had an omega coefficient below .80. The mean inter-item correlation within each scale is also higher than the longer versions, with the exception of the Attention Shifting scale which was likely due to an increased scale length. Improved reliability reflects reduced measurement error which, from a CTT perspective, indicates observed scores more accurately represent true scores (DeVellis, 2017) and therefore more precisely measure the constructs of interest.

Furthermore, the revised scales demonstrate superior CFA properties than the original, indicating greater construct validity and better utility for structural equation modeling (Brown, 2006). Modification indices of the unstructured CFA models suggest a stronger need for covaried residuals among items of the original scales compared to the revised scales. This stems from the original scales including item pairs with very similar wording (e.g., Inhibitory Control Item 6 / CBQ 32 *"My child has a hard time following instructions"* and Item 7 / CBQ 136 *"My child is good at following instructions"*, M.I. = 13.80). Although using reverse-worded items is an easy way to improve mean inter-item correlation and subsequently α, large modification indices for covarying residuals among these item pairs proves a strong relation exists between the items beyond that which the factor explains. In a CFA, we can accommodate this relation by covarying item residuals. In CTT, however, this relationship is not modeled, and therefore the observed score is a more biased estimate of the true score. These item pairs also overweight the representation of the item content by essentially counting a single item twice when computing the scale score. By avoiding item pairs that need covaried residuals within CFA, the revised scales provide more accurate measurement of the latent constructs in both CTT and unstructured CFA frameworks.

The revised scale content and definitions also suggest improved measurement accuracy over the original, as new items were written and final items selected based on scale definitions, and the scale definitions were modified to better reflect content. Changes to the Inhibitory Control scale highlight this point. We decided early on to omit "planning" from the Inhibitory Control definition, as it was perceived to not be a core component of the construct. This altered the definition and scale content by removing from consideration CBQ 64 (*"My child prepares for trips and outings by planning things s/he will need."*). We initially chose to include aspects of care and caution in the scale definition and item pool, although after examining the pilot data and reviewing the scale with Dr. Rothbart, we chose to omit these aspects as they may be confounded by fear. I would argue that the revised Inhibitory Control scale is truer to the hypothesized construct, as it only incorporates content related to the core nature of inhibitory control—the ability to inhibit or delay an action—and not secondary aspects or confounded items.

The addition of an Activation Control scale also is an improvement over the original CBQ. Activation Control is an EC construct typically measured only in late childhood and beyond using questionnaires such as the TMCQ (Simonds & Rothbart, 2004) and the Early Adolescence Temperament Questionnaire (EATQ; Ellis & Rothbart, 2001) or in lab tasks such as the Mistaken Gift Paradigm (Simonds, Kieras, Rueda, & Rothbart, 2007). While no literature suggests Activation Control does not emerge until age seven, there is some evidence that children's ability to activate a subdominant response increases with age. For instance, Simonds et al. (2007) found that between ages 7 and 10, children became better able smile when presented with a disappointing present. A literature search for studies using parent- or self-report measures of Activation Control shows that most studies use the scale only as an indicator of Effortful Control (e.g., Muris, Meesters, & Blijlevens, 2007; Valiente et al., 2008), so based on work with older children it isunclear what utility a measure of Activation Control for younger children will provide. However, a number of developmental studies with younger children examine fine-grained aspects of temperament (e.g., Jones, Rothbart, & Posner, 2003; Putnam, Rothbart, & Gartstein, 2008; Rudasill, Gallagher, & White, 2010), and therefore a scale that explicitly measures Activation Control in children ages three to seven may be welcome.

The revised scales should also allow for better investigations into the development of EC, by reducing longitudinal measurement noninvariance and therefore bias in scale scores. While in the current study I was unable to establish true longitudinal measurement invariance by measuring EC in children at age three and again at age seven—I would like to graduate, after all—multigroup invariance testing by age is the closest I could come to establishing longitudinal invariance within a cross-sectional design. This design does not measure intra-individual differences but still helps identify items that perform differently due to the child age. For example, Item 217 (*"My child has a hard time waiting his/her turn to talk when excited."*) was more strongly related to the latent factor of Inhibitory Control for the older group of children than the younger group (metric noninvariance). This indicates that the item is more salient to Inhibitory Control for older children than younger children. When a scale shows measurement noninvariance, it would be inappropriate to interpret group differences or longitudinal changes, as at least some of the difference may be due to properties of the measurement instrument and not necessarily the construct of interest (Brown, 2006). Thus, the revised scales can give researchers confidence in measuring fine-grained aspects of effortful control over time.

The revised scales also represent a step toward allowing for more consistent measurement of temperament beyond ages three and seven. First, the updated rating scale ensures the revised scales are measured with the same anchors as earlier measures of temperament (i.e., ITQ and ECBQ). Therefore, parent ratings at least represent EC constructs perceived in the same way (i.e., as frequency of observed behaviors). Second, I made an explicit attempt to incorporate questions from the ECBQ and TMCQ into the revised scales, in the hopes of finding some items that may be measurement invariant within this age range and may be used to equate scale scores of the CBQ EC scales and those of these other measures. Two ECBQ items were retained in the revised scales (Item 9 / ECBQ 189 *"When I'm busy, my child can find another activity to do when asked."*, Activation Control for the revised scale, Attention Shifting for ECBQ scale; Item 111 /

ECBQ 157 *"If spoken to when watching TV, my child will reply appropriately then go back to watching TV."*, Attention Shifting), and one TMCQ item was retained as well (Item 358 / TMCQ 123 *"My child likes to run his/her hand over things to see if they are smooth or rough."*, Perceptual Sensitivity). Although not all revised scales contained items from temperament measures for younger and older children, the overlap that does occur provides some evidence of construct validity across these ages.

Other common measures of temperament have been developed or updated relatively recently (Gartstein & Rothbart, 2003; Putnam, Gartstein, & Rothbart, 2006; Simonds & Rothbart, 2004), whereas the CBQ is essentially unchanged for the past thirty years. Like the IBQ, ECBQ, and TMCQ, the revised version of the CBQ EC scales represent a more current understanding of temperament. For instance, during theoretical refinement of the final scales, Dr. Rothbart noted that care and caution are conceptually confounded with fear, and therefore should be omitted from the Inhibitory Control scales. We also chose to further refine Inhibitory Control by removing planning from the definition and item content. Examining the ECBQ and CBQ-R Inhibitory Control scales, the content overlap is much better clearer, with predominantly aspects of inhibiting behavior and delay of gratification represented. In this example, while the revised Inhibitory Control scale does not contain any ECBQ items, we can still be confident that the constructs measured are similar given content overlap. This should, however, be tested with longitudinal data in the future to confirm.

Putnam, Rothbart, and Gartstein (2008) investigated longitudinal relations among temperament constructs measured in early childhood with the ECBQ, and later in childhood with the CBQ. Their approach, in which they correlated ECBQ scale scores at about age two with CBQ scale scores at about age four, showed moderate stability between constructs. For instance, Attentional Focusing, Low intensity Pleasure, and Perceptual Sensitivity had similar stability ($r = .40$, $r = .40$, $r = .41$, respectively, $p < .05$), while Inhibitory Control was a bit higher ($r = .51$, $p < .05$). Given the updated scale content and the generally improved psychometric properties of the revised scales, I would expect a similar study correlating scores of the ECBQ and CBQ-R EC scales would show higher coefficients, due to reduced measurement error and better aligned content.

Finally, the revised CBQ EC scales present an advantage over the original because new items were written to apply more consistently across diverse contexts (e.g., culture, SES). One of the selection criteria for considering revised items was whether the item would hold the same meaning in different settings. Dr. Carmen González Salinas was instrumental in helping identify items that could not be conveyed in the same manner when translated into Spanish, which otherwise appeared fine from the other content experts in the United States. For instance, Carmen flagged the proposed Inhibitory Control item *“When I’m in another room and my child wants to get my attention, s/he will call out or yell for me instead of coming into the room to talk to me.”* because she noted this behavior was not necessarily unacceptable to families in Spain, and may be confounded with socioeconomic status (i.e., it’s more commonplace among low SES families than high SES families). It should be noted, however, that while we tried to avoid items that may be confounded by culture, there may still be some cultures in which certain items may perform poorly. For instance, several items in the revised scales contain references to television. For researchers using the revised CBQ EC scales with families with limited access to television (e.g., a developing country), it would be advisable to adjust the wording of these items to refer to a similarly engaging, and more culturally salient, activity.

**Disadvantages of the Revised CBQ Effortful Control Scales**

Although the CBQ-R Low Intensity Pleasure scale has somewhat better psychometric properties than the original—better unconstrained CFA fit, poorer CTT reliability—the question of the scale's utility remains. In our personal correspondence documented in Chapter 3, Dr. Ahadi suggested that Low Intensity Pleasure may require both active and inactive aspects (e.g., completing a puzzle vs. taking a warm bath), and thus recommended we try to incorporate these types of items in the rewritten scale. Indeed, we wrote several "active" items involving cognitive engagement in low intensity activities, but none of them remained in the final scale due to low loadings, low variability, or noninvariance. Thus, the remaining scale appears to be a scale of "passive" Low Intensity Pleasure. Rueda (2012) notes that EC reflects the voluntary aspects of control, as opposed to involuntary aspects such as anxiety and fear. In this sense, it could be argued that the original and current CBQ-R Low Intensity Pleasures scales do not represent voluntary aspects of control, but rather an involuntary proclivity toward calm, quiet activities. This is further evidenced by its relatively low bivariate correlations with the EC scales of Activation Control, Attentional Focusing, Attention Shifting, and Inhibitory Control—these values range between $r$ = .26 and $r$ = .45, compared to correlations among those other scales ranging from $r$ = .56 and $r$ = .78 (see Table 3.13). Yet with this line of reasoning, the Perceptual Sensitivity scale may also be suspect as an indicator of EC. Perceptual Sensitivity has the lowest inter-scale correlation, ranging from a non-significant $r$ = .12 with Inhibitory Control to $r$ = .34 with Attentional

Focusing and $r = .42$ with Low Intensity Pleasure. Furthermore, these two scales do not show demonstrable age-group differences (see Table 3.21).

Perhaps Low Intensity Pleasure and Perceptual Sensitivity should not be incorporated as CBQ-R measures of EC. In many, if not most, empirical studies of EC in children, the construct is represented as some combination of Attentional Focusing, Attention Shifting, and Inhibitory Control with total disregard to the Low Intensity Pleasure and Perceptual Sensitivity Scales (e.g., Blair & Razza, 2007; Diaz et al., 2016; Eisenberg et al., 2007; Sanchez-Perez et al., 2015; Valiente et al., 2003). The general disregard for these two scales is an indication that the broader scientific community interested in EC do not consider Low Intensity Pleasure or Perceptual Sensitivity as salient aspects of this factor.

An important aspect of scale development is the consideration of how the scale will be used (DeVellis, 2017). If most researchers conceptualize EC as only involving attentional and inhibitory aspects, does this study presents sufficient evidence to *demand* the inclusion of Low Intensity Pleasure and Perceptual Sensitivity in the measurement of EC? No, probably not. But, after all, the purpose of this project was to ensure adequate measurement of the fine-grained aspects that Rothbart and colleagues have conceptualized as components of effortful control. I did not seek to establish the factor structure of EC, although evidence provided here shows expected relations between these aspects, to varying degrees. Personally, I would argue that the Low Intensity Pleasure scale reflects secondary aspects of EC, and that this scale is subject to age-related measurement error, as many of its items appear to decrease with age (e.g., Item 255 / CBQ 174 *"My child enjoys sitting on a parent's lap."*). I am less inclined to disregard the *construct* of Perceptual Sensitivity as an element of EC, but rather feel its measurement via parent report is far more subjective than the other aspects of EC and therefore is limited in its relation to the overall factor. For instance, Item 380 (*"My child notices small changes in the environment, like lights getting brighter in a room."*) requires parents to be especially attuned to their child's own sensitivity to subtle environmental stimuli. This scale may also be confounded by verbal ability or surgency, as children who vocalize their detection of minor changes in the environment would likely be rated higher than children who detect the same changes but do not acknowledge them verbally. I would recommend to researchers interested in using the CBQ-R to at least consider including all EC scales, but drop Low Intensity Pleasure and/or Perceptual Sensitivity if needed.

The CBQ is one of the most well-known and utilized instruments for measuring children's temperament for research purposes. Given its ubiquity, one disadvantage of revising the CBQ scales is the potential uncertainty a revised version casts on prior research using the instrument. A search of the PsychInfo database indicates the original Rothbart, Ahadi, Hershey, and Fisher (2001) article which formally provided reliability and validity evidence for the CBQ has been cited 729 times (as of 6/6/2017). Furthermore, as the CBQ was used prior to 2001, there are likely thousands of studies that have used the CBQ in one form or the other. Age-group noninvariance identified in Study 1 of this dissertation, as well as issues of CBQ factor structure identified elsewhere (e.g., Kotelnikova et al., 2016), suggest that results of previous studies using the CBQ that were contrary to hypotheses or conflicted with other findings may be in part due to measurement issues with the EC scales. Researchers currently using the CBQ scales I suspect the new version would not likely invalidate any previous studies of child temperament, at least where temperament was measured at a single time-point, primarily because the original and revised versions correlate so highly and overlap in content. With the sample in Study 2, the original and revised scales correlate highly—between $r = .800$ and $r = .900$ for the scales, and $r = .930$ for effortful control overall. I would think we can be reasonably confident that correlational findings between the original CBQ EC scales and socioemotional, academic, or other outcomes would likely still hold. However, there is no way of truly knowing this without replication.

Regarding longitudinal research, the original CBQ scales clearly show some degree of measurement noninvariance within a CFA context. However, I could find no studies to date that have examined longitudinal change in individual dimensions of children's effortful control using the CBQ within a CFA context; most longitudinal investigations use scale arithmetic means. For instance, Moilanen and colleagues (2009) examined predictors of longitudinal change in parent reported inhibitory control from age two to age four. Using latent growth curve modeling, the authors found a linear increase over this period in inhibitory control, measured at each time point as the arithmetic mean of the scale items. Within the CTT framework, the unit of analysis is the total score, not the item, and the difficulty and discrimination of individual items are weighted the same in computation (Edwards & Wirth, 2009). Therefore, the authors could only assume measurement invariance because there is no way to test for it when using a CTT-produced manifest variable. A more accurate way for examining longitudinal change in this context would be to estimate CFA measurement models at each time point with Inhibitory Control as the latent variable and empirically test for invariance. Once invariance is established, we could be confident observed changes are a function of change over time rather than measurement error.

Another disadvantage of the revised CBQ scales is the focus on parent reported child temperament. There has long been evidence of discrepancies between parent-ratings and ratings from other informants (e.g., teachers, observers) of child temperament using a variety of measures (e.g., Allan et al., 2013; Billman & McDevitt, 1980; Victor, Halverson, & Wampler, 1988), with some evidence that teacher-report of children's EC may be favorable over parent report as these ratings align more closely with behavioral EC tasks (Allan et al., 2014). Teglasi et al., (2015) modified the CBQ-SF for teachers, and found significant correlations between parent- and teacher-reported Attentional Focusing and Inhibitory Control in Kindergarten ($r$ = .45 & $r$ = .63, $p$s < .01), but not for Low Intensity Pleasure ($r$ = -.03, $p$ > .05) or Perceptual Sensitivity ($r$ = .01, $p$ > .05). Discrepancies are not unique to the study of EC, as low correspondence has been found between parent- and teacher ratings of other temperament dimensions, such as Shyness (Rudasill et al., 2014). There are a number of reasons parents and other reporters may differ in their ratings of children's temperament, and Rothbart and Bates (2006) provide a detailed overview of the sources of error in the measurement of temperament. One reason parents and teachers may differ in their ratings is the reference group against which they may compare a given child's behavior. Parents may not have much experience caring for children other than their own, limiting their knowledge of the typicality of certain behaviors. Teachers, on the other hand, have a wealth of experience with many children and thus hold a broad knowledge of typical behaviors for children of a certain age. However, the depth of knowledge teachers hold about individual children may vary and may certainly be different than that of parents. Parent reports are useful, of course, as parents have much broader knowledge of the behavior of their child across different contexts and ages, and parents have been shown to be valid reporters of child temperament (see: Rothbart & Bates, 2006), yet it may still be prudent to adjust for bias in parent report. This point will be discussed in the next section of this chapter, Limitations and Future Directions.

One unfortunate byproduct of introducing a revised version of the CBQ EC scales now relates to the proliferation of translations and psychometric evaluations of the CBQ over the past several years. At least a dozen papers have been published recently examining the factor structure, reliability, and validity evidence of the CBQ, CBQ-SF, or CBQ-VSF (Allan et al., 2013; Backer-Grøndahl et al., 2016; Clark et al., 2016; de la Osa et al., 2014; Kotelnikova et al., 2016; Leyfer et al., 2012; Richard, Davis, & Burns, 2008; Roberts et al., 2014; Sleddens et al., 2011 Sleddens et al., 2012; Sulik et al, 2010). The CBQ standard form has also been translated in at least 15 languages (http://research.bowdoin.edu/). By revising the CBQ EC scales—and potentially the remaining scales—these exercises in understanding the measurement properties of the original CBQ, applying the CBQ to special populations (e.g., Leyfer et al., 2012; Richard et al., 2008; Roberts et al., 2014), and translating the standard version for use in other countries (e.g., Sleddens et al., 2011), may need to be revisited. Longitudinal studies and large-scale studies using the original CBQ are potentially at a disadvantage as well, because once the revised scales are published, the resulting longitudinal datasets will contain parent reports of temperament using the original version of the instrument with identified and established flaws. This is an issue inherent in longitudinal research, and researchers are faced with either introducing the updated scale midway through the study, or retaining the original. This is a decision for the investigators to make based on their study aims and research design.

## Limitations and Future Directions

A number of limitations of Study 2 should be noted. The primary limitation is that the sample is comprised of mainly white, middle- to upper-class mothers and their children living in urbanized areas in the United States with access to the internet and Facebook. Fathers, and parents of other ethnicities, countries of residence, and lower socioeconomic backgrounds were not well represented as reporters in both the calibration and validation samples, and therefore the sample used to create the scale is not diverse enough to be certain the psychometric properties of the revised scales reported here would match those for the greater population with a balanced proportion of mothers and fathers, and a broader range of ethnic and SES backgrounds. Clark et al. (2016) found the Low Intensity Pleasure scale demonstrated scalar noninvariance based on parent (mother vs. father report), so it is possible that revising the scales based on a sample with an even balance between father- and mother-report would have produced slightly different scales. Sulik et al. (2010) used the CBQ Attentional Focusing and Inhibitory Control scales a two of several indicators of EC, including some behavioral tasks, and found measurement invariance across ethnicity (White vs. Black vs. Hispanic). Future studies should investigate how psychometric properties of the revised scales compare based on parent gender and ethnicity.

Since such a homogenous sample was used for the selection of items, and the scales validated using a similar sample, we can only be confident these scales are valid for the population represented here. Because I used internal consistency to somewhat guide item selection with the homogeneous calibration sample, the estimates of internal consistency with the similar validation sample may be higher than found in future studies using more diverse samples. The same could be said about scale factor structures. However, this limitation is not critical for several reasons First, items dropped during calibration for having too many “N/A” responses, low variability, and low item-total correlations and factor loadings are still poor items for measuring effortful control, namely because they do not perform well for the current sample. Even if these dropped items were to perform well with other segments of the population, they do not work with married white women, and therefore should be dropped anyway. The items dropped based on the CFA decision rules involving covaried residuals and noninvariance may be candidates to replace items that perform poorly in other samples, yet the items dropped based on content expert feedback should not be considered because they failed theoretical evaluation. Second, it is possible that a diverse sample would provide more variability in item responses and therefore may *better* discriminate among children and increase internal consistency, although the opposite is also plausible. Third, other studies have shown the original CBQ generalizes well to other populations, such as low-income families (Richard et al., 2008), so there is hope the revised version may be generalizable as well. Future studies are needed to establish the CBQ-R EC scales as a valid, reliable measures of effortful control for children of low-SES families, racial and ethnic minorities, and men.

Another limitation is that the current study only approximates longitudinal measurement invariance by using a cross-sectional design to establish multi-group invariance by age. True longitudinal measurement invariance would need to be tested in a longitudinal design and provide effect sizes of measurement invariance, such as the mean and covariance structure (MACS) analysis presented by Nye and Drasgow (2011). There is some indication that the original CBQ EC scales demonstrate measurement noninvariance at the factor level by both child gender and parent report (mother vs. father), mainly due to the Low Intensity Pleasure scale (Clark et al., 2016). Clark and colleagues (2016) present a clear and concise application of Mean and Covariance Structure Analysis (MACS) applied to CBQ data for estimating the effect of measurement noninvariance, and this procedure could be usefully applied to longitudinal data. While it was not within the scope of the present study to test for other forms of measurement invariance, future studies should evaluate the revised EC scales to verify the constructs are measured consistently regardless of age, child and parent gender, reporter, ethnicity, and SES.

It should be noted, however, that Clark et al (2016) examined measurement invariance at the factor level, whereas the current study examined measurement invariance at the scale level. Longitudinal measurement noninvariance at the factor level (i.e., how the scales relate to the EC factor at different ages) may well be found, but this may not reflect failings of the individual scales, but rather developmental variability in the composition of EC. Assuming the EC scales themselves demonstrate longitudinal invariance, findings of factor-level longitudinal *noninvariance* would suggest the structure of EC changes over time. The purpose of the present study was not to investigate the factor structure of EC, as has been done so many times before, but rather to ensure fine-grained measurement of effortful control constructs could be consistently measured over time. With this in mind, it might be useful for the study of temperament and child development to examine longitudinal changes in the factor structure of the CBQ-R EC scales. Such an undertaking may also help to clarify the nature of Low Intensity Pleasure and Perceptual Sensitivity in relation to EC, as (a) secondary aspects of the construct, (b) tangential attributes hamstrung by measurement error, (c) core components, or (d) unrelated to EC.

The current revision of the CBQ also focused solely on the EC scales. There are still two factors—Negative Affectivity and Extraversion/Surgency—comprised of 11 other scales that could perhaps use revision. While I and the content experts have no immediate plans to revise the remaining scales, conversations with contributors to this project suggest this is the next step. Such an undertaking should follow the same procedure as Study 1 to identify problematic items and scales, and Study 2 to revise and evaluate the scales. Furthermore, some studies have included EFAs of CBQ data and found factor structures different than Rothbart and colleagues (2001) reported (e.g., Kotelnikova et al., 2016). Thus, it would be useful to include the revised EC scales in the data collection to not only evaluate revised Negative Affectivity and Extraversion/Surgency scales, but to examine the factor structure of items, and perhaps identify items that may be confounded with other constructs.

As mentioned in the previous section, parent report of temperament may be somewhat biased, given the reference group against which they compare their child’s behavior. One recent innovation, anchoring vignettes, may provide a means for correcting for discrepancies between parent reported measures of temperament and measures from other sources. King and Wand (2007) presented the use of anchoring vignettes to address differential item functioning (DIF) in survey responses. Using this technique, participants are presented a vignette that briefly describes a fictional person, and asks the respondent to rate them on a scale of interest. For example, an anchoring vignette to correct parent responses for inhibitory control would be administered at the end of the questionnaire, and may look like this:

> *"It's Sally's birthday today, and three wrapped presents are sitting on a table in the dining room, waiting to be opened. When her mother is not looking, Sally leaves her toys, slowly approaches the table and quietly examines the presents. She lifts one of the boxes to feel its weight. She peeks inside a bag, trying to get a glimpse of the present beneath the tissue paper. Sally then hears her mother returning from the kitchen. Fearing she will get caught snooping around her presents, Sally runs from the table and returns to her toys.*
>
> *"On a scale of 1 to 7, how would you rate Sally's ability to delay gratification and control inappropriate behaviors?"*

From here, the participant's response to the Inhibitory Control items would be rescaled to be relative to their assessment of Sally's level of inhibitory control. Assuming that the participants respond to the questionnaire items and the vignette in the same way, and that the vignette is interpreted in the same way by all respondents, King and Wand (2007) argue that anchoring vignettes adequately correct for DIF and put responses on a common scale by anchoring a reference point. Other studies have provided support for anchoring vignettes with self-report measures (e.g., Soest et al., 2011), yet I am not aware of any studies to have employed them using parent report. This would be a useful innovation to address the problem reference group bias noted in parent reports.

It is also common to use the CBQ in its various forms for teacher reported child temperament (Allan et al., 2013; Rudasill et al., 2014; Teglasi et al., 2015), although Rothbart and colleagues (2001) did not explicitly design the CBQ to be used with teachers as informants. Teacher reports, like parent reports, are not without bias. While teachers may have a wealth of experience with many students against which they can rate the temperament of a single child, they have only limited experience with each student and limited contexts within which to observe students. A few studies use slight modifications to make the CBQ more applicable to the contexts in which teachers observe children (i.e., Allan et al., 2013; Teglasi et al., 2015), yet there is no indication these modifications are consistent across studies. Like the original version, several items on the revised scales may refer to situations or contexts teachers are unable to observe (e.g., Item 310 / CBQ 54 *"My child enjoys taking warm baths."*; Item 216 *"When I'm on a phone call and my child wants to talk to me, s/he will wait until I've finished the call."*; Item 106 *"My child has difficulty turning off the TV when told to do so."*). Rather than leaving it up to individual research groups to determine how the revised CBQ scales should be modified, it would be worthwhile to create and validate a teacher-report version in the future.

One of the main purposes of this project was to ensure EC was measured consistently across ages three to seven. However, developmental researchers have long been interested in the structure and development of temperament, and particularly EC, over longer spans of time (e.g., Carranza, González-Salinas, & Ato, 2013; Putnam, Ellis,

& Rothbart, 2001; Putnam, Rothbart, & Gartstein, 2008). Yet these studies are restricted to the scales available and the ages at which they are useful. For example, the ECBQ has been validated for children age 18 months to 36 months, while the CBQ has been validated for children age three to age seven. When using longitudinal designs that study children across the age three threshold, researchers are faced with either using one scale at ages outside the validated age range for repeated measures (e.g., Moilanen et al., 2009), or using different scales at different ages. Neither is ideal. Using an age inappropriate measure is questionable without validity evidence supporting its use, such as verifying the scale holds the same factor structure as in the intended population. Using a separate measure presents further questions of content and construct equivalence. Potential solutions to this dilemma would be to either establish the validity of existing scales at these age ranges, or create hybrid scales that span an age range across this threshold. Toward addressing the latter point, in the present study I purposefully considered ECBQ and TMCQ items in hopes of identifying invariant items that could span a larger age range. While few items from these scales were retained, this represents a step in such a direction. Further work is needed using children of all ages and items from multiple scales to address such a complex issue.

Finally, while the aspects of EC should certainly overlap, future studies are needed to verify the revised CBQ EC scales are measuring distinct constructs. From Table 3.13, we see that the revised Activation Control, Attentional Focusing, Attention Shifting, and Inhibitory Control scales correlate highly ($r$ = .56 to $r$ = .78). Activation Control and Inhibitory Control are the two scales that appear to be most similar both statistically ($r$ = .78) and conceptually. As an illustration, imagine a child playing with a ball in their front yard and the ball gets kicked across the street. The child runs to the edge of the street, looks both ways, then crosses the street after ensuring no moving vehicles are nearby. The child did not immediately run into the street to fetch their ball—suppressing a dominant response—and therefore displayed inhibitory control. However, the child also stopped at the side of the street and looked both ways—performing a subdominant response—and therefore also displayed activation control. There are a few ways of establishing evidence of discriminant validity. First, an exploratory factor analysis such as PCA—with oblique rotation, as the factors should be correlated—could be used at the item level to determine if the loading patterns match the scale compositions. Second, each scale should correlate most highly with a different criterion designed to measure the same construct, and correlate less-so with other criterion measuring different constructs. For example, child engagement during the bead-sorting task from the Laboratory Temperament Assessment Battery (LAB-TAB; Goldsmith & Rothbart, 1999) is purported to measure attentional focusing, whereas the task Grass/Snow—where a child points to a white square when the experimenter says “Grass” and points to a green square when the experimenter says “Snow” (Carlson & Moses, 2001)—measures inhibitory control. Therefore, the revised CBQ Attentional Focusing scale should correlate higher with the bead sorting task than Grass/Snow, and the revised Inhibitory Control scale should correlate higher with Grass/Snow than bead sorting. The Low Intensity Pleasure and Perceptual Sensitivity scales correlate only modestly with the other four scales ($r$ = .26 to $r$ = .45 and $r$ = .12 and $r$ = .34, respectively), so there seems less of a need establish the distinctiveness of these two scales between the others.

## Conclusions

The revised CBQ EC scales show considerable improvement upon the original scales. With the help of content experts, the scales and their content areas were clearly defined, and the revised scales arguably represent content of the hypothesized constructs better than the original. Poor items identified in Study 1 were mostly omitted from the revised scales, as were those that may be confounded with other dimensions of temperament and items tapping secondary aspects of effortful control.

While the revised and original scales correlate highly, the revised scales have demonstrably better psychometric properties than the original. With the exception of the Attention Shifting and Attentional Focusing scales, all scales were reduced in length while maintaining or even improving internal consistency and mean inter-item correlations. The Attention Shifting scale was extended in length, and demonstrated a slightly higher internal consistency than the original scales, whereas the revised Attentional Focusing scale remained at 9 items, but with superior psychometric properties. When placed into unconstrained CFAs, fit of the revised scales was markedly better than original scales. The revised scales appear to be unidimensional, while several of the original scales have some indication of multidimensionality. Compared to the original scales, the revised scales showed considerably less age-group measurement noninvariance. Finally, original and revised scales were comparably stable based on scores one-month apart. Thus, the revised scales appear to produce reliable and valid measurement of EC in children ages three to seven. While the revised scales probably will not discredit prior work on EC, they will allow for better investigations in to the development of EC over time and may facilitate a broader understanding of EC through the addition of the Activation Control scale. Because of the emphasis on flagging culturally problematic items during the item writing process, the revised scales may also permit a consistent measure of temperament across cultures and geographic areas, yet further research is needed to establish validity evidence to support this point.

The revisions presented here are not without drawbacks. The revised scales were devised as a parent report measure, and therefore they still suffer from biases associated with this source. Employing anchoring vignettes or retooling and validating these scales for use with teachers may help overcome some bias. In addition, the samples used to calibrate and validate the scales were rather homogenous, and thus further work to demonstrate the utility of these scales based on a broader sample is needed. Finally, the Negative Affectivity and Extraversion/Surgency scales should also be evaluated for measurement invariance and revised as needed. The present study provides a good outline for how to accomplish this, and these results suggest such an endeavor would be worthwhile.

# REFERENCES


Allan, N. P, Lonigan, C. J., & Wilson, S. B. (2013) Psychometric evaluation of the Children's Behavior Questionnaire-Very Short Form in preschool children using parent and teacher report. *Early childhood research quarterly, 28*. 302-313.

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental psychology, 50*(10), 2368-2379.

Backer-Grøndahl, A., Nærde, A., Ulleberg, P., & Janson, H. (2016). Measuring effortful control using the Children's Behavior Questionnaire—Very Short Form: Modeling matters. *Journal of personality assessment, 98*(1), 100-109.

Bandura, A. (2005) The primacy of self-regulation in health promotion. *Applied psychology: An international review 54*(2), 245-254.

Bell, M. A., & Deater-Deckard, K. (2007). Biological systems and the development of self-regulation: Integrating behavior, genetics, and psychophysiology. *Journal of developmental & behavioral pediatrics*, *28*(5), 409-420.

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: differential susceptibility to environmental influences. *Psychological bulletin, 135*(6), 885–908.

Bentler, P. M. (1990) Comparative fit indexes in structural models. *Psychological bulletin, 107*(2) 238-246.

Billman, J. & McDevitt, S. C. (1980). Convergence of parent and observer ratings of temperament with observations of peer interaction in nursery school. *Child development, 51*, 395-400.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American psychologist*, *57*(2), 111-127.

Blair, C. & Razza, R. P. (2007) Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, *78*(2), 647-663.

Bridgett, D. J., Oddi, K. B., Laake, L. M., Murdock, K. W., & Bachmann, M. N. (2013). Integrating and differentiating aspects of self-regulation: Effortful control, executive functioning, and links to negative affectivity. *Emotion*, *13*(1), 47-63.

Brown, T. A. (2006), *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage focus editions*, *154*, 136-136.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, *4*(6), 215-222.

Calkins, S. D. (2007). The emergence of self-regulation: Biological and behavioral control mechanisms supporting toddler competencies. In C. A. Brownell, & C. B. Kopp (Eds.) *Socioemotional development in the toddler years: Transitions and transformations*, 261-284. NY: Guilford Press

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental neuropsychology*, *28*(2), 595-616.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children’s theory of mind. *Child development, 72*, 1032–1053.

Carranza, J. A., González-Salinas, C., & Ato, E. (2013). A longitudinal study of temperament continuity through the IBQ, TBAQ, and CBQ. *Infant behavior and development, 36*, 749-761.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research, 1*, 245-276.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, *9*(2), 233-255.

Clark, D. A., Listro, C. J., Lo, S. L., Durbin, C. E., Donnellan, M. B., & Neppl, T. K. (2016). Measurement Invariance and Child Temperament: An Evaluation of Sex and Informant Differences on the Child Behavior Questionnaire. *Psychological assessment*, 1-17.

Cortina, J. M. (1993). What is coefficient α? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98-104.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological bulletin*, *115*(1), 74.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. New York, NY: Guilford Publications.

de la Osa, N., Granero, R., Penelo, E., Domènech, J. M., & Ezpeleta, L. (2013). The short and very short forms of the children's behavior questionnaire in a community sample of preschoolers. *Assessment, 21*(4) 463–476.

Derryberry, D., & Reed, M. A. (1994). Temperament and the self-organization of personality. *Development and psychopathology, 6*, 653-676. 18

Derryberry, D., & Reed, M. A. (1996). Regulatory processes and the development of cognitive representations. *Development and psychopathology, 8*, 215-234.

Derryberry, D., & Rothbart, M. K. (1988). Arousal, affect, and attention as components of temperament. *Journal of personality and social psychology*, *55*(6), 958-966.

DeVellis, R. F. (2012). *Scale development: Theory and applications (3rd ed.).* Thousand Oaks, CA: SAGE Publications.

DeVellis, R. F. (2017). *Scale development: Theory and applications (4th ed.).* Los Angeles: SAGE Publications.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey, & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 67-110). Hillsdale, NJ: Erlbaum.

Diaz, A., Berger, R., Valiente, C., Eisenberg, N., VanSchyndel, S. K., Tao, C., Spinrad, T., Doane, L. D., Thompson, M. S., Silva, K. M., & Southworth, J. (2016). Children's sleep and academic achievement: The moderating role of effortful control. *International journal of behavioral development, 42*(2), 275-284.

Duncan, O. D. (1984). *Notes on social measurement: Historical and critical.* Russell Sage Foundation.

Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From α to ω: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, *105*, 399-412.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of life research*, *16*(1), 5-18.

Edwards, M. C. & Wirth, R. J. (2009). Measurement and the study of change. *Research in human development*, *6*(2-3), 74-96.

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., & Guthrie, I. K. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child development*, *72*(4), 1112-1134.

Eisenberg, N., Michalik, N., Spinrad, T. L., Hofer, C., Kupfer, A., Valiente, C., Liew, J., Cumberland, A., & Reiser, M. (2007). The relations of effortful control and impulsivity to children's sympathy: A longitudinal study. *Cognitive development*, *22*(4), 544-567.

Eisenberg, N., Valiente, C., & Eggum, N. D. (2010). Self-regulation and school readiness. *Early education and development*, *21*(5), 681-698.

Eisenberg, N., Valiente, C., Fabes, R. A., Smith, C. L., Reiser, M., Shepard, S. A., Losoya, S. H., Guthrie, I. K., Murphy, B. C. & Cumberland, A. J. (2003). The relations of effortful control and ego control to children's resiliency and social functioning. *Developmental psychology*, *39*(4), 761-776.

Ellis, L. K., & Rothbart, M. K. (2001). *Revision of the Early Adolescent Temperament Questionnaire.* Poster presented at the 2001 Biennial Meeting of the Society for Research in Child Development, Minneapolis, Minnesota.

Fiske, D. W. (1971). *Measuring the concepts of personality.* Chicago: Aldine.

Frohn, S., Prokasky, A. A., & Putnam, S. (October, 2016). *Reconstructing the CBQ-VSF Effortful Control Scale: Comparing CTT and IRT Approaches*. Poster presented at the Occasional Temperament Conference in Seattle, WA.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant behavior and development*, *26*(1), 64-86.

Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24-36 months of age. *Developmental Science, 3/4,* 397-404.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of consumer research*, *11*(1), 572-580.

Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In J. Strelau & A. Angleitner (Eds.) *Explorations in temperament*, (pp. 249-272). New York: Springer.

Goldsmith, H. H., & Rothbart, M. K. (1999). *The Laboratory Temperament Assessment Battery (Locomotor Version Edition 3.1)*. Madison, WI: University of Wisconsin–Madison.

González, C., Carranza, J. A., Fuentes, L. J., Galián, M. D., & Estévez, A. F. (2001). Attentional mechanisms and development of self-regulation in childhood. *Anales de psicologia, 17*(2), 275-286.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161.

Hambleton, R. K. & Jones, R. W. (1993) An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational measurement: issues and practice*, *12*(3), 38-47.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child development*, *72*(2), 625-638.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods, 3*(4), 424-453.

Hubley, A. M., & Zumbo, B. D. (2013). Psychometric characteristics of assessment procedures: An overview. In K. F. Geisinger (Ed.) *APA Handbook of Testing and Assessment in Psychology Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology* (pp. 1-19). APA.

Iyer, R. V., Kochenderfer-Ladd, B., Eisenberg, N., & Thompson, M. (2010). Peer victimization and effortful control: Relations to school engagement and academic achievement. *Merrill-Palmer quarterly (Wayne State University. Press), 56*(3), 361-386.

Jones, L., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental science, 6*, 498-504.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, *20*, 141-151.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486-507.

Kim, S., Nordling, J. K., Yoon, J. E., Boldt, L. J., & Kochanska, G. (2013). Effortful control in "hot" and "cool" tasks differentially predicts children's behavior problems and academic performance. *Journal of abnormal child psychology*, *41*(1), 43-56.

King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political analysis, 15*, 46-66.

Kingston, N. M., Scheuring, S. T., Kramer, L. B. (2013). Test development strategies. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 165-184). Washington, DC: APA.

Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., & Grey, M. (2007). The analysis and interpretation of cognitive interviews for instrument development. *Research in nursing & health*, *30*(2), 224-234.

Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child development*, *72*(4), 1091-1111.

Kotelnikova, Y., Olino, T. M., Klein, D. N., Mackrell, S. V., & Hayden, E. P. (2016). Higher and Lower Order Factor Analyses of the Temperament in Middle Childhood Questionnaire. *Assessment*, *28*(1) 92–108.

Lengua, L. J., Bush, N. R., Long, A. C., Kovacs, E. A., & Trancik, A. M. (2008). Effortful control as a moderator of the relation between contextual risk factors and growth in adjustment problems. *Development and psychopathology*, *20*(02), 509-528.

Leyfer, O., John, A. E., Woodruff-Borden, J., & Mervis, C. B. (2012). Factor structure of the Children's Behavior Questionnaire in children with Williams syndrome. *Journal of autism and developmental disorders, 42*(11), 2346-2353.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, *1*(2), 130-149.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of applied psychology, 93,* 568–592.

Moilanen, K. L., Shaw, D. S., Dishion, T. J., Gardner, F., & Wilson, M. (2010). Predictors of longitudinal growth in inhibitory control in early childhood. *Social development, 19*(2), 326-347.

Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In *Handbook of structural equation modeling* (pp. 380–393). New York, NY: Guilford Press.

Muris, P., Meesters, C., & Blijlevens, P. (2007). Self-reported reactive and regulative temperament in early adolescence: Relations to internalizing and externalizing problem behavior and "Big Three" personality factors. *Journal of adolescence, 30*(6), 1035-1049.

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus User's Guide (6$^{th}$ ed).* Los Angeles, CA: Muthén & Muthén.

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of applied psychology, 96,* 966–980.

Paulussen-Hoogeboom, M. C., Stams, G. J. J., Hermanns, J., & Peetsma, T. T. (2007). Child negative emotionality and parenting from infancy to preschool: A meta-analytic review. *Developmental Psychology*, *43*(2), 438-453.

Penfield. (2013).Item analysis. In K. F. Geisinger (Ed.) *APA Handbook of Testing and Assessment in Psychology Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology* (pp. 121-138). APA.

Pluess, M., & Belsky, J. (2010). Differential susceptibility to parenting and quality child care. *Developmental Psychology*, *46*(2), 379–90.

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and psychopathology*, *12*(03), 427-441.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1-23.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, *104*(1), 1-15.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of personality assessment*, *87*(1), 102-112.

Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasz & A. Angelinter (Eds.) *Advances in* *research on temperament*, 165-182. Lengerich, Germany: Pabst Science Publishers.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant behavior and development*, *29*(3), 386-401.

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and child development*, *17*(4), 387-405.

Putnam, S. P., Sanson, A. V., & Rothbart, M. K. (2002). Child temperament and parenting. *Handbook of parenting*, *1*, 255-277.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raiche, G. (2010). *nFactors: An R package for parallel analysis and non graphical solutions to the Cattell scree test*. R package version 2.3.3.

Raiche, G., Riopel, M. & Blais, J. G. (2006). *Non graphical solutions for the Cattell's scree test.* Paper presented at the International Annual meeting of the Psychometric Society, Montreal.

Revelle, W. (2017) *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA. R package version 1.7.3.

Rhoades, B. L., Greenberg, M. T., & Domitrovich, C. E. (2009). The contribution of inhibitory control to preschoolers' social–emotional competence. *Journal of applied developmental psychology*, *30*(3), 310-320.

Richard, H., Davis, D. W., & Burns, B. M. (2008). An evaluation of the Children's Behavior Questionnaire for use with children from low-income families. *Journal of early childhood and infant psychology, 4,* 111.

Roberts, J. E., Tonnsen, B. L., Robinson, M., McQuillin, S. D., & Hatton, D. D. (2014). Temperament factor structure in fragile X syndrome: The Children's Behavior Questionnaire. *Research in developmental disabilities*, *35*(2), 563-571.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child development*, 569-578.

Rothbart, M. K. (1982). The concept of difficult temperament: A critical analysis of Thomas, Chess, and Korn. *Merrill-Palmer quarterly 28*(1), 35-40.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current directions in psychological science, 16,* 207–212.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. (6th ed.) In N. Eisenberg (Ed.), *Social, emotional, and personality development, Vol. 3,* (pp. 99–166). New York, NY: Wiley.

Rothbart, M.K., Derryberry, D., and Posner, M.I. (1994). A psychobiological approach to the development of temperament. In J.E. Bates & T.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 83-116). Washington, DC: American Psychological Association.

Rothbart, M. K., & Jones, L. B. (1998). Temperament, self-regulation, and education. *School psychology review*, *27*(4), 479-491.

Rothbart, M. K., & Rueda, M. R. (2005). The development of effortful control. In U. Mayr, E. Awh, & S. W. Keele (Eds.) *Developing individuality in the human brain: A tribute to Michael I. Posner* (167-188). APA.

Rothbart, M. K., & Rueda, M. R. (2005). The development of effortful control. *Developing individuality in the human brain: A tribute to Michael I. Posner*, 167-188.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child development*, *72*(5), 1394-1408.

Rothbart, M. K., Ellis, L. K., Rosario Rueda, M., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of personality*, *71*(6), 1113-1144.

Rothbart, M. K., Sheese, B. E., & Posner, M. I. (2007). Executive attention and effortful control: Linking temperament, brain networks, and genes. *Child development perspectives*, *1*(1), 2-7.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer quarterly*, *40*(1), 21–39.

Rudasill, K. M., Gallagher, K. C., & White, J. M. (2010). Temperamental attention and activity, classroom emotional support, and academic achievement in third grade. *Journal of school psychology*, *48*(2), 113–34.

Rudasill, K. M., Prokasky, A., Tu, X., Frohn, S., Sirota, K., & Molfese, V. J. (2014). Parent and teacher ratings of children's shyness as predictors of early academic skills. *Learning and individual differences*, *34*, 57-62

Rueda, M. R. (2012). Effortful control. In M. Zentner & Shiner, R. L (Eds.) *Handbook of temperament* (pp. 145-167). New York: Guilford.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029-1040.

Samejima, F. (2010). The general graded response model. In M. L. Nering & R. Ostini (Eds), *Handbook of polytomous item response theory models*, (pp. 43-76). New York NY: Routledge.

Sánchez-Pérez, N., Fuentes, L. J., Pina, V., López-López, J. A., & González-Salinas, (2016). How do different components of effortful control contribute to children’s mathematics achievement? *Frontiers in psychology*, *6*, 1-15.

Simonds, J., & Rothbart, M. K. (2004, October). The Temperament in Middle Childhood Questionnaire (TMCQ): A computerized self-report instrument for ages 7–10. *Poster presented at the biennial Occasional Temperament Conference*. Athens, GA.

Sireci, S. G. & Sukin, T. (2013). Test validity. In K. F. Geisinger (Ed.) *APA Handbook of Testing and Assessment in Psychology Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology* (pp. 62-85). APA.

Sleddens, E. F., Hughes, S. O., O'Connor, T. M., Beltran, A., Baranowski, J. C., Nicklas, T. A., & Baranowski, T. (2012). The Children's Behavior Questionnaire very short scale: psychometric properties and development of a one-item temperament scale. *Psychological reports, 110*(1), 197-217.

Sleddens, E. F., Kremers, S. P., Candel, M. J., De Vries, N. N., & Thijs, C. (2011). Validating the Children's Behavior Questionnaire in Dutch children: Psychometric properties and a cross-cultural comparison of factor structures. *Psychological assessment*, *23*(2), 417.

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the royal statistical society: Series A (statistics in society)*, *174*(3), 575-595.

Stevens, S. S. (1946). On the theory of measurement scales. *Science*, *103*, 677-680.

Stright, A. D., Gallagher, K. C., & Kelley, K. (2008). Infant temperament moderates relations between maternal parenting in early childhood and children's adjustment in first grade. *Child development*, *79*(1), 186-200.

Sulik, M. J., Huerta, S., Zerr, A. A., Eisenberg, N., Spinrad, T. L., Valiente, C., Di Giunta, L., Pina, A. A., Eggum, N. D., Sallquist, J., Edwards, A., Kupfer, A., Lonigan, C. J., Phillips, B. M., Wilson, S. B., Clancy-Menchetti, J., Landry, S. H., Swank, P. R., Assel, M. A., & Taylor, H. B. (2010). The factor structure of effortful control and measurement invariance across ethnicity and sex in a high-risk sample. *Journal of psychopathology and behavioral assessment*, *32*(1), 8-22.

Teglasi, H., Schussler, L., Gifford, K., Annotti, L. A., Sanders, C., & Liu, H. (2015). Child behavior questionnaire–Short form for teachers: Informant correspondences and divergences. *Assessment*, *22*(6), 730-748.

Thomas, A., & Chess, S. (1977). *Temperament and development*. New York: University Press.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with" well fitting" models. *Journal of abnormal psychology*, *112*(4), 578-598.

Valiente, C., Eisenberg, N., Smith, C. L., Reiser, M., Fabes, R. A., Losoya, S., Guthrie, I. K., & Murphy, B. C. (2003). The relations of effortful control and reactive control to children's externalizing problems: A longitudinal assessment. *Journal of personality*, *71*(6), 1171-1196.

Valiente, C., Lemery-Chalfant, K., Swanson, J., & Reiser, M. (2008). Prediction of children's academic competence from their effortful control, relationships, and classroom participation. *Journal of educational psychology*, *100*(1), 67-77.

Vermeulen, M. C., Astill, R. G., Benjamins, J. S., Swaab, H., Van Someren, E. J., & Van der Heijden, K. B. (2016). Temperament moderates the association between sleep duration and cognitive performance in children. *Journal of experimental child psychology*, *144*, 184-198.

Victor, J. B., Halverson, C. F., & Wampler, K. S. (1988). Family-school context: Parent and teacher agreement on child temperament. *Journal of consulting and clinical psychology*, *56*(4), 573.

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, *18*(1), 53-70.

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.

Willoughby, M., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behavior and academic achievement. *Developmental neuropsychology*, *36*(2), 162-180.

Wolfe, V. V., Finch, A. J., Saylor, C. F., Blount, R. L., Pallmeyer, T. P., & Carek, D. J. (1987). Negative affectivity in children: A multitrait-multimethod investigation. *Journal of consulting and clinical psychology*, *55*(2), 245-250.

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: a call for an integrated model of self-regulation. *Child development perspectives*, *6*(2), 112-121.

Zhou, Q., Eisenberg, N., Wang, Y., & Reiser, M. (2004). Chinese children's effortful control and dispositional anger/frustration: relations to parenting styles and children's social functioning. *Developmental psychology*, *40*(3), 352-366.

APPENDIX A: PAPER-FORM CBQ-R EFFORTFUL CONTROL QUESTIONNAIRE

**Children's Behavior Questionnaire – Revised**
Effortful Control Subscales

Participant ID: ___________________ Child Gender: _____________________

Today's Date: ____________________ (mm/dd/yyyy) Child Birthdate: ____________________ (mm/dd/yyyy)

Child Age: ________ _________
years months

**Instructions**

On the next pages you will see a set of statements that describe children's reactions in a number of situations. We would like you to tell us what your child's reaction has been in those situations. There are no "correct" ways of reacting; children differ widely in their reactions, and it is these differences we are trying to learn about.

As you read each statement about your child's behavior, please indicate how often your child did this during the last month by using the following rating scale:

| Never | Very Rarely | Less than Half the Time | Half the Time | More than Half the Time | Almost Always | Always | Does not Apply |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |

**Note about "does not apply" versus "Never"**

If you cannot answer one of the items because you have never seen the child in that situation, then select does not apply. For example, if the statement is about the child's reaction to your singing and you have never sung to your child, then select does not apply.

However, if you have never noticed your child behave in a described way, then select Never. For example, if the statement is about your child touching objects to feel their texture, and you have never noticed your child do this, then select Never.

**Please be sure to select a number or NA for every item.**

| Never | Very Rarely | Less than Half the Time | Half the Time | More than Half the Time | Almost Always | Always | Does not Apply |
|---|---|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **NA** |

Thinking about the past **month**…

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | My child takes care of needed tasks (e.g., washing hands, going to the bathroom, cleaning up) before playing. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 2 | My child notices when very quiet music is playing in the background. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 3 | My child is easily distracted when listening to a story. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 4 | My child can focus attention promptly when cued, for example attending to directions. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 5 | My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 6 | When practicing an activity, my child has a hard time keeping her/his mind on it. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 7 | When picking up toys or doing other jobs, my child usually keeps at the task until it's done. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 8 | My child will stay with the same activity for a half-hour or more if given the chance. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 9 | My child can easily stop a play activity when told to stop. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 10 | When around flowers, my child will take the time to smell them and enjoy their scent. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 11 | When involved in an activity, my child answers quickly if I call him/her. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 12 | My child likes the sound of words, as in nursery rhymes or poems. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 13 | My child has a hard time following instructions. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 14 | My child quickly gives up on an activity or project when s/he becomes bored of it. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |

| Never | Very Rarely | Less than Half the Time | Half the Time | More than Half the Time | Almost Always | Always | Does not Apply |
|---|---|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **NA** |

Thinking about the past **month**…

| # | Item | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 | If spoken to when watching TV, my child will reply appropriately then go back to watching TV. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 16 | My child enjoys sitting on a parent's lap. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 17 | My child sometimes doesn't seem to hear me when I talk to her/him. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 18 | My child has an easy time waiting to open a present. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 19 | My child seems to listen to even quiet sounds. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 20 | My child will participate in a play activity, even if it is not the activity they would prefer. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 21 | My child likes to run his/her hand over things to see if they are smooth or rough. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 22 | It is hard to get my child's attention when s/he is concentrating on something. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 23 | When needing to walk long distances, my child will keep walking even when feeling tired. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 24 | My child spends a long time engaged in drawing, coloring, or crafts. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 25 | My child seeks opportunities to get hugs and kisses from family members. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 26 | My child gets easily distracted when drawing, reading, or playing alone. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 27 | My child seems to notice when a room is more warm or cool than expected. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 28 | My child is usually able to resist temptation when told s/he is not supposed to do something. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 29 | My child notices small changes in the environment, like lights getting brighter in a room. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |

| Never | Very Rarely | Less than Half the Time | Half the Time | More than Half the Time | Almost Always | Always | Does not Apply |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |

Thinking about the past **month**…

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 30 | It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 31 | My child can only pay attention to one activity or source of information at a time. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 32 | My child likes it when a parent or caregiver sings to him/her. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 33 | My child comments when a parent has changed his/her appearance. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 34 | My child has an easy time leaving play to come to dinner. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 35 | My child has a hard time waiting for his/her favorite snack or meal. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 36 | When trying to learn how a new toy works, my child concentrates intensely. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 37 | My child enjoys gentle rhythmic activities, such as rocking or swaying. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 38 | My child seems to notice when sounds change in volume, such as when someone turns the radio or television up or down. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 39 | My child has difficulty turning off the TV when told to do so. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 40 | My child follows rules for games, rather than making up his/her own rules to suit them. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 41 | My child enjoys listening to music quietly. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 42 | My child has trouble sitting still when s/he is told to (at movies, church, etc.). | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 43 | When I'm busy, my child can find another activity to do when asked. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 44 | My child pays close attention when something is being explained to him or her. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |

| Never | Very Rarely | Less than Half the Time | Half the Time | More than Half the Time | Almost Always | Always | Does not Apply |
|---|---|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **NA** |

Thinking about the past **month**…

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 45 | My child enjoys taking warm baths. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 46 | My child continues to focus on what they are doing, even if there are other things going on around them. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 47 | My child will notice different aromas in nature (e.g., flowers), even when the scents are subtle. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 48 | When I'm on a phone call and my child wants to talk to me, s/he will wait until I've finished the call. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 49 | My child sometimes becomes absorbed in a picture book and looks at it for a long time. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 50 | My child enjoys just being talked to. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 51 | My child can lower his/her voice when asked to do so. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |
| 52 | When asked to look at something, my child does so immediately. | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | NA |

# APPENDIX B: PAPER-FORM CBQ-R EFFORTFUL CONTROL SCORESHEET

## Children's Behavior Questionnaire – Revised

Effortful Control Subscales Scoresheet

Scale scores for the Children's Behavior Questionnaire – Revised represent the mean score of all scale items applicable to the child during the last month, as judged by the caregiver. Scales scores are to be computed by the following method:

1. Sum all numerical item responses for a given scale. Note that:
    a. If caregiver omitted an item, that item receives no numerical score;
    b. If caregiver checked the "does not apply" response option for an item, that item receives no numerical score;
    c. Items indicated with an (R) are reverse items and must be scored in the following way:

| | |
|---|---|
| 7 becomes 1 | 3 becomes 5 |
| 6 becomes 2 | 2 becomes 6 |
| 5 becomes 3 | 1 becomes 7 |
| 4 remains 4 | |

2. Divide the total by the number of items receiving a numerical response. Do not include items marked "does not apply" or items receiving no response in determining the number of items.

Effortful Control – Whole Scale: 52 items, standardized α ≈ .92

Activation Control: 10 items, standardized α ≈ .82

*The capacity to perform a subdominant action against a natural inclination or aversion.*

1 My child takes care of needed tasks (e.g., washing hands, going to the bathroom, cleaning up) before playing.

5 My child will persist at tasks that demand concentration (e.g., complex puzzles, projects), even if s/he has to struggle to complete them.

7 When picking up toys or doing other jobs, my child usually keeps at the task until it's done.

14 (R) My child quickly gives up on an activity or project when s/he becomes bored of it.

20 My child will participate in a play activity, even if it is not the activity they would prefer.

23 When needing to walk long distances, my child will keep walking even when feeling tired.

30 (R) It takes a long time to get my child engaged in an activity they are not interested in, such as chores or homework.

34 My child has an easy time leaving play to come to dinner.

40 My child follows rules for games, rather than making up his/her own rules to suit them.

43 When I'm busy, my child can find another activity to do when asked.

Attentional Focusing: 9 items, standardized α ≈ .80

*The capacity to maintain attentional focus upon task-related channels.*

4 My child can focus attention promptly when cued, for example attending to directions.

11 When involved in an activity, my child answers quickly if I call him/her.

15 If spoken to when watching TV, my child will reply appropriately then go back to watching TV.

17 (R) My child sometimes doesn't seem to hear me when I talk to her/him.

22 It is hard to get my child's attention when s/he is concentrating on something.

31 (R) My child can only pay attention to one activity or source of information at a time.

39 (R) My child has difficulty turning off the TV when told to do so.

52 When asked to look at something, my child does so immediately.

<u>Attention Shifting</u>: 8 items, standardized α ≈ .84
*The capacity to voluntarily shift attention from one activity to another.*

| | | |
|---|---|---|
| 3 | (R) | My child is easily distracted when listening to a story. |
| 6 | (R) | When practicing an activity, my child has a hard time keeping her/his mind on it. |
| 8 | | My child will stay with the same activity for a half-hour or more if given the chance. |
| 24 | | My child spends a long time engaged in drawing, coloring, or crafts. |
| 26 | (R) | My child gets easily distracted when drawing, reading, or playing alone. |
| 36 | | When trying to learn how a new toy works, my child concentrates intensely. |
| 44 | | My child pays close attention when something is being explained to him or her. |
| 46 | | My child continues to focus on what they are doing, even if there are other things going on around them. |
| 49 | | My child sometimes becomes absorbed in a picture book and looks at it for a long time. |

<u>Inhibitory Control</u>: 8 items, standardized α ≈ .85
*The capacity to suppress or moderate desired behaviors and delay actions.*

| | | |
|---|---|---|
| 9 | | My child can easily stop a play activity when told to stop. |
| 13 | (R) | My child has a hard time following instructions. |
| 18 | | My child has an easy time waiting to open a present. |
| 28 | | My child is usually able to resist temptation when told s/he is not supposed to do something. |
| 35 | (R) | My child has a hard time waiting for his/her favorite snack or meal. |
| 42 | (R) | My child has trouble sitting still when s/he is told to (at movies, church, etc.). |
| 48 | | When I'm on a phone call and my child wants to talk to me, s/he will wait until I've finished the call. |
| 51 | | My child can lower his/her voice when asked to do so. |

Low Intensity Pleasure: 9 items, standardized α ≈ .73
*Amount of pleasure or enjoyment related to situations involving low stimulus intensity, rate, complexity, novelty, and incongruity.*

10 When around flowers, my child will take the time to smell them and enjoy their scent.
12 My child likes the sound of words, as in nursery rhymes or poems.
16 My child enjoys sitting on a parent's lap.
25 My child seeks opportunities to get hugs and kisses from family members.
32 My child likes it when a parent or caregiver sings to him/her.
37 My child enjoys gentle rhythmic activities, such as rocking or swaying.
41 My child enjoys listening to music quietly.
45 My child enjoys taking warm baths.
50 My child enjoys just being talked to.

Perceptual Sensitivity: 8 items, standardized α ≈ .82
*Amount of detection of slight, low intensity stimuli from the external environment.*

2 My child notices when very quiet music is playing in the background.
19 My child seems to listen to even quiet sounds.
21 My child likes to run his/her hand over things to see if they are smooth or rough.
27 My child seems to notice when a room is more warm or cool than expected.
29 My child notices small changes in the environment, like lights getting brighter in a room.
33 My child comments when a parent has changed his/her appearance.
38 My child seems to notice when sounds change in volume, such as when someone turns the radio or television up or down.
47 My child will notice different aromas in nature (e.g., flowers), even when the scents are subtle.